

A. Markowski, J. Kostro, A. Lewandowski

AUTOMATYKA

w pytaniach
i odpowiedziach

AUTOMATYKA

w pytaniach
i odpowiedziach

Wydawnictwa Naukowo-Techniczne
Warszawa 1979

Opiniodawca
doc. dr hab. inż. Konstanty Jan Kurman
Redaktor naukowy WNT
mgr inż. Maria Frankowska
Redaktor techniczny
Anna Napiórkowska
Okładkę projektował
Andrzej Stroka

62.5

W książce przedstawiono w przystępny sposób, w formie pytań i odpowiedzi, najważniejsze pojęcia automatyki. Podano zasady opisu właściwości obiektów statycznych i dynamicznych, liniowych i nieliniowych, ciągłych i dyskretnych. Wyjaśniono podstawy regulacji automatycznej obiektów, ich analizy i projektowania. Omówiono technikę systemów automatyzacji i budowę urządzeń automatyki oraz zastosowanie urządzeń cyfrowych w automatyce. Wyjaśniono pojęcia optymalizacji oraz zasady działania i zastosowania maszyn analogowych i hybrydowych. Książka jest przeznaczona dla szerokiego kręgu Czytelników — inżynierów i techników różnych specjalności oraz dla studentów i uczniów szkół technicznych w zakresie podstaw automatyki.

© Copyright by Wydawnictwa Naukowo-Techniczne,
Warszawa 1979

All rights reserved

Printed in Poland

ISBN 83-204-0110-0

WNT Warszawa 1979 Wydanie I. Nakład 20.000+260 egz. Ark. wyd. 20,1 Ark. druk. 16,25 (21,61/A) Format B5. Papier offset, kl. III 80 g. 70×100. Pracę oddano do składania 28.XII 1978 r. Podpisano do druku w listopadzie 1979 r. Druk ukończono w grudniu 1979 r. Cena zł 56.— Symbol ET/79267/WNT
Drukarnia im. Rew. Październikowej. Zam. 45/79 C-38

Spis treści

Rozdział 1

Wiadomości wstępne (7-11)

Automatyka · Sterowanie · Regulacja · Układ regulacji

Rozdział 2

Opis i właściwości obiektów automatyzacji (12-77)

Model obiektu · Zmienne · Stan układu · Układ statyczny, dynamiczny · Układy liniowe · Układy statyczne nieliniowe · Linearyzacja · Połączenia układów · Układy dynamiczne liniowe ciągłe · Opis za pomocą równań różniczkowych · Transmitancje · Charakterystyki częstotliwościowe i czasowe · Całka spłotowa · Wartości własne · Macierz podstawowa · Stabilność · Właściwości dynamiczne typowych układów · Zera transmitancji · Opis stanowy przy danej transmitancji · Układy niestacjonarne · Układy dyskretne · Układy o parametrach rozłożonych · Układy nieliniowe · Nieliniowość statyki · Płaszczyzna fazowa · Punkty osobliwe

Rozdział 3

Podstawy regulacji automatycznej (78-110)

Struktura układu regulacji · Zadania regulacji · Klasyfikacja układów regulacji · Wskaźniki jakości regulacji · Stabilność układu regulacji, metoda Nyquista, metoda linii pierwiastkowych, metoda funkcji opisującej · Projektowanie układów regulacji · Korekcja charakterystyk częstotliwościowych · Regulator PID · Serwomechanizmy, serwomechanizm liniowy, serwomechanizm przekątnikowy

Rozdział 4

Technika systemów automatyzacji (111-158)

Systemy aparatury regulacyjnej · Elementy nastawcze · Siłowniki · Czujniki · Przetworniki pomiarowe ·

Regulatory, regulatory PID, regulatory pneumatyczne
PID, regulatory hydrauliczne • Rejestratory • Dobór
nastaw regulatorów • Poprawa działania układu
regulacji • Regulacja kaskadowa

Rozdział 5

Urządzenia cyfrowe w układach automatyki (159-185)

Zadania realizowane przez urządzenia cyfrowe •
Urządzenia centralnej rejestracji i przetwarzania danych •
Bezpośrednie sterowanie cyfrowe • Elementy urządzeń
cyfrowych • Elementy logiczne • Komputerowy system
sterowania • Połączenie obiektów przemysłowych
z urządzeniami cyfrowymi • Przetworniki analogowo-
cyfrowe • Przetworniki cyfrowo-analogowe •
Niezwadność urządzeń cyfrowych • Serwomechanizm
cyfrowy

Rozdział 6

Optymalizacja i jej zastosowanie w automatyce (186-236)

Optymalność • Wskaźnik jakości • Decyzje dopuszczalne •
Programowanie matematyczne • Istnienie rozwiązań
zadania optymalizacji • Zadania wariacyjne • Zadania
z ograniczeniami • Zadania wariacyjne z ograniczeniami •
Numeryczne rozwiązywanie zadań optymalizacji •
Zastosowania optymalizacji w automatyce •
Optymalizacja parametryczna • Regulacja ekstremalna •
Regulacja adaptacyjna • Identyfikacja obiektu •
Regulator optymalny • Zasada optymalności

Rozdział 7

Maszyny matematyczne i ich zastosowanie (237-260)

Symulacja • Zasada analogii • Maszyna analogowa •
Wzmacniacz operacyjny • Układy realizowane przy
użyciu wzmacniacza operacyjnego • Programowanie
maszyny analogowej • Rozwiązywanie równań
różniczkowych • Zastosowanie do syntezy układu
regulacji • Integrator sterowany • Maszyna iteracyjna •
Programowanie maszyny iteracyjnej • Porównanie
maszyny cyfrowej i analogowej • Maszyna hybrydowa

·1.1· Co to jest automatyka?

Automatyka jest to dziedzina wiedzy, która zajmuje się możliwościami ograniczenia lub wyeliminowania udziału człowieka w czynnościach związanych ze sterowaniem różnorodnych obiektów fizycznych.

Terminem „automatyka” określa się też potocznie zestaw (system) urządzeń technicznych (aparatów) przeznaczonych do samoczynnego sterowania np. procesu przemysłowego.

W określeniu automatyki istotne jest podkreślenie samoczynności (braku lub ograniczenia udziału obsługi ludzkiej); tak więc „automatycznie” jest przeciwieństwem do „ręcznie”, zwłaszcza w odniesieniu do pojęcia sterowania.

·1.2· Co to jest automatyzacja?

Przez automatyzację rozumie się działalność natury technicznej, gospodarczej i organizacyjnej, zmierzającą do wprowadzenia metod i środków automatyki w rozmaite dziedziny życia. Automatyzacji podlega w pierwszym rzędzie produkcja przemysłowa — jest to przejaw dążenia do osiągnięcia możliwie dużej wydajności przy jednoczesnym zapewnieniu lepszych i bezpieczniejszych warunków pracy. Automatyzacja obejmuje w coraz większym stopniu transport, łączność, administrację (szczególne znaczenie ma automatyzacja prac rachunkowych oraz przetwarzania informacji przy użyciu maszyn cyfrowych). Trzeba też wymienić ochronę zdrowia, szkolnictwo, gospodarstwo domowe — jako kolejne działy, w których stosuje się automatyzację. Ważnym polem zastosowań automatyki jest obronność kraju. Bez automatyzacji nie byłby możliwy rozwój badań kosmicznych. Istnieje potoczne i uzasadnione przekonanie, że automatyzacja jest jedną z głównych cech nowoczesności i postępu technicznego.

·1.3· Co to jest sterowanie?

Sterowaniem nazywa się oddziaływanie na dany obiekt, w sposób zamierzony, mający doprowadzić do spełnienia określonego celu. Sterowanie z zasady nie łączy się bezpo-

średnio z wydatkiem energii, lecz najczęściej jest wyrażone przez sygnał reprezentujący pewien strumień informacji. Natomiast efektem sterowania mogą być zmiany przepływu znacznych strumieni energii. Obiekt, na który wywiera się oddziaływanie sterujące, nazywa się **obiektem sterowanym**.

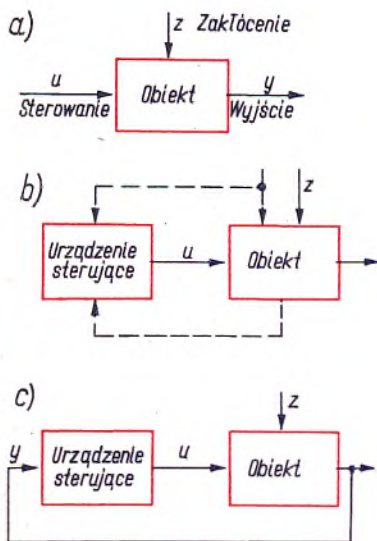
Na przykład piec opalany gazem może stanowić obiekt sterowany, przy czym sterowanie polega na zmianach ustawienia zaworu w przewodzie doprowadzającym gaz. Sterowanie dopuszczalne mieści się oczywiście między całkowitym zamknięciem zaworu a jego pełnym otwarciem. Cel sterowania może polegać na ustaleniu określonej temperatury w piecu przez odpowiednie ustawienie zaworu. Zauważmy, że na tak określony cel wpływa nie tylko oddziaływanie uznane przez nas za sterowanie — temperatura w piecu zależy bowiem również od zawartości pieca, dopływu powietrza, ciśnienia gazu przed zaworem i wielu innych czynników.

Czynniki o charakterze przypadkowym, niezamierzonym, niekontrolowanym, utrudniające sterowanie, nazywa się **zakłóceniami**.

1.4. Czym różni się sterowanie od regulacji?

Istnieją dwa zasadnicze sposoby sterowania: sterowanie w **układzie otwartym** i sterowanie w **układzie zamkniętym** (czyli w układzie ze **sprężeniem zwrotnym**). Regulacja oznacza drugi z wymienionych sposobów, jest więc pojęciem węższym od pojęcia sterowania. Regulacja, czyli sterowanie ze sprzężeniem zwrotnym, jest jednym z najważniejszych pojęć automatyki, dlatego wyjaśnimy je dokładniej (rys. 1.4). Przedstawimy symbolicznie za pomocą bloczka obiekt sterowany (rys. 1.4a), na który oddziałuje wielkość sterująca (sygnał sterujący) u i zakłócenie z , i przypuścmy, że cel sterowania można określić za pośrednictwem pewnej cechy wielkości wyjściowej (sterowanej) obiektu — y , np. cel ten może oznaczać ustalenie określonej wartości wielkości y .

Zacznijmy od sterowania w układzie otwartym. Na rysunku 1.4b zaznaczono obok obiektu urządzenie sterujące tj. człon, w którym wytwarza się sygnał u reprezentujący sterowanie obiektu. Układ taki nazywa się **układem otwartym sterowania**. Urządzenie sterujące nie otrzymuje tu żadnych informacji o właściwej wielkości regulowanej y , lecz co najwyżej pewne dane dotyczące pożądanego celu sterowania. Ponadto urządzenie sterujące może otrzymywać pewne informacje o zakłóceniach lub też o wielkościach pomocniczych, charakteryzujących pracę obiektu. Jak widać, sterowanie w układzie otwartym ma sens wtedy, gdy na podstawie znajomości sygnału sterującego można dokładnie przewidzieć przebieg wielkości regulowanej. W najlepszym razie przewidywanie takie może być ułatwione, jeśli mamy dostęp do pomiaru zakłóceń lub wielkości pomocniczych, jak to zaznaczono na rysunku. Jeśli jednak pojawi się zakłócenie, którego wpływu nie znamy, to sterowanie w układzie otwartym będzie



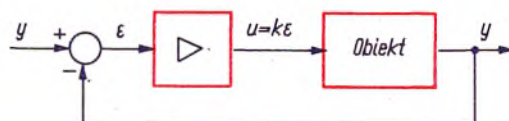
1.4. Sterowanie i regulacja: a) symbol obiektu sterowanego i sygnałów: sterującego, zakłócającego i wyjściowego; b) układ otwarty sterowania; c) układ zamknięty sterowania (układ regulacji)

nieskuteczne. W przykładzie obiektu z poprzedniego pytania sterowanie w układzie otwartym polegałoby, przy żądaniu utrzymania określonej stałej temperatury w piecu, na ustawieniu zaworu dopływu gazu w pewnym położeniu, takim, o którym wiedzielibyśmy z góry, że powinno odpowiadać żądanej temperaturze; wystarczy jednak, że wystąpi jakiegokolwiek zakłócenie, a ustawione położenie zaworu może okazać się niewłaściwe.

Aby otrzymać **układ zamknięty sterowania**, czyli układ regulacji, należy zamknąć pętlę oddziaływań, tzn. uzależnić sterowanie od skutków, jakie to sterowanie wywołuje, co symbolicznie pokazano na rys. 1.4c. Połączenie wielkości regulowanej y , zamykające pętlę regulacji, nazywa się **sprzężeniem zwrotnym**. Wprowadzenie sygnału reprezentującego wielkość regulowaną do urządzenia sterującego uniezależnia nas od zakłóceń, ponieważ kontrola skutków sterowania umożliwia bieżące korygowanie tego sterowania. Jest to podstawowa zaleta regulacji — dążenie do osiągnięcia właściwego celu, bez względu na oddziaływanie zakłóceń. W tym celu potrzebne jest porównywanie postawionego celu z aktualnie obserwowanym wynikiem sterowania i przystosowywanie tego sterowania do ewentualnie dostrzeżonych odchyleń realizacji celu. Wracając do naszego przykładu, układ regulacji otrzymany przez zainstalowanie termometru w piecu i uzależnienie ustawienia zaworu od wskazań termometru: jeśli temperatura jest za niska, należy zawór bardziej otworzyć, jeśli za wysoka — przyknać. Zauważmy, że regulacja nie musi być automatyczna — jeśli manipulacji zaworem dokonuje człowiek patrząc na termometr, to nazwiemy to regulacją ręczną. Oczywiście nasze zainteresowanie będzie skierowane na **regulację automatyczną**, kiedy opisane operacje są dokonywane samoczynnie przez regulator, czyli urządzenie sterujące obiektem na podstawie danych pomiarowych pobieranych z odpowiednich czujników zainstalowanych w obiekcie. Regulator powinien również otrzymać sygnał reprezentujący cel sterowania — np. wartość temperatury, która ma być utrzymywana w piecu.

1.5. Jaki jest najprostszy układ regulacji automatycznej?

Najprostszy układ regulacji, którego celem jest utrzymywanie **wartości zadanej** wielkości regulowanej, pokazano na rys. 1.5. Sygnał reprezentujący wartość zadaną y_0 jest dany z zewnątrz. Porównując wartość zadaną z aktualną wartością sygnału regulowanego, czyli dokonując odejmowania (na rysunku zaznaczono to symbolicznie za pomocą kółka ze strzałkami odpowiednich sygnałów i znakami), otrzymuje się tzw. **uchyby regulacji** $\varepsilon = y_0 - y$. Jeśli teraz uzależnimy

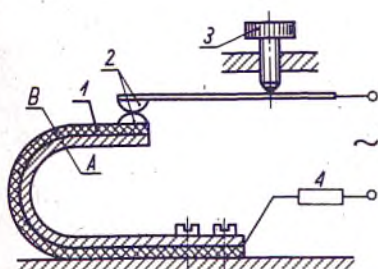


1.5. Schemat blokowy najprostszego układu regulacji

sterowanie obiektu u od uchybu ε w taki sposób, że w przypadku uchybu dodatniego, tzn. $\varepsilon > 0$ czyli $y_0 > y$, wywołamy sterowanie zwiększające y , zaś w przypadku przeciwnym, $\varepsilon < 0$ i $y_0 < y$, wywołamy zmniejszanie y , to zawsze uzyskamy tendencję zmierzającą do zmniejszenia uchybu (możliwie do zera) czyli zbliżenia (możliwie zrównania) wartości y_0 i y . W tym celu czasem wystarczy uzależnienie sygnału sterującego u od sygnału ε na zasadzie zwykłego wzmacnienia — tak jak zaznaczono na rysunku — jeśli tylko dodatnie wartości u powodują zwiększanie wartości y . Stąd wniosek, że najprostszym regulatorem jest odpowiednio dobrany wzmacniacz sygnału ε .

·1.6· Z jakim najprostszym układem regulacji automatycznej można się spotkać w życiu codziennym?

Każdy z nas zapewne miał w ręku żelazko elektryczne z tzw. termoregulatorem. Zawiera ono bardzo prosty, ale skuteczny układ regulacji temperatury. Zwykłe żelazko nagrzewa się wtedy, kiedy jest włączone do sieci elektrycznej, zaś stygnie, kiedy jest wyłączone. Zarówno nagrzewanie, jak stygnięcie przebiegają dość wolno. Jeśli wyłączymy nagrzane żelazko po przekroczeniu żądanej temperatury, to dopiero po chwili temperatura wyraźnie maleje. Jeśli wtedy żelazko ponownie włączymy, to znowu po chwili temperatura przekroczy żądany poziom itd. Przy takim działaniu średnia temperatura cały czas jest bliska żądanej wartości. Do uzyskania automatycznej realizacji opisanego sposobu działania jest potrzebny element, który odłącza dopływ prądu przy zbyt wysokiej temperaturze i włącza po obniżeniu temperatury. W omawianym przypadku elementem tym jest tzw. bimetaliczny czujnik temperatury, połączony z układem stykowym zapewniającym zamykanie lub przerywanie obwodu prądu. Schemat takiego termoregulatora pokazano na rys. 1.6. Czujnik bimetaliczny jest to sprężyna wykonana z dwóch złączonych pasków metali o różnych współczynnikach rozszerzalności cieplnej. Przy wzroście temperatury następuje wyginanie sprężyny w stronę o mniejszej rozszerzalności. Jak widać z rysunku, sprężyna bimetaliczna z umieszczonym na końcu stykiem w stanie spoczynkowym zamyka obwód prądu przepływającego przez grzejnik żelazka. Jeśli temperatura dostatecznie wzrośnie, sprężyna wygnie się tak dalece, że styki rozewrą się, obwód prądu zostanie przerwany, grzejnik przestanie wydłazić ciepło i temperatura zacznie maleć. Po zmniejszeniu temperatury sprężyna powróci do położenia spoczynkowego, obwód grzejnika zostanie zamknięty, temperatura zacznie wzrastać, po czym sytuacja będzie się powtarzać. Patrząc na ten prosty przykład układu regulacji zauważmy, że sprężyna bimetaliczna jest tu elementem pomiarowym (jej wygięcie odwzorowuje zmiany temperatury) i jednocześnie regulacyjnym (styk umieszczony na sprężynie współpracując ze stykiem nieruchomym zamyka i otwiera obwód prądu, a więc steruje proces nagrze-



·1.6· Schemat działania termoregulatora bimetalicznego

1 — sprężyna bimetaliczna (A — strona o mniejszej rozszerzalności, B — o większej); 2 — styki;
3 — pokrętko wartości zadanej temperatury;
4 — grzejnik

wania lub stygnięcia). Stan pracy (nagrzewanie lub stygnięcie) sygnalizuje zwykle lampka. Pozostaje jeszcze ustalić sposób wprowadzania wartości zadanej. Realizuje się to przez ustawienie wstępnego napięcia sprężyny tj. docisku styków w stanie spoczynkowym, co zapewnia pokrętło umieszczone na żelazku. Przy niewielkim napięciu wstępnym wystarczy stosunkowo niska temperatura dla rozwarcia styków. Przy dużym napięciu wstępnym styki są mocno dociśnięte i dopiero po znacznym wzroście temperatury następuje ich rozwarcie.

Podobne proste układy regulacyjne stosuje się w urządzeniach domowych i w mniej odpowiedzialnych układach regulacji (głównie temperatury) w przemyśle.

·2.1· Co to jest obiekt automatyzacji?

Przez obiekt będziemy rozumieć układ fizyczny lub proces technologiczny, będący przedmiotem automatyzacji. W szerokim sensie obiektem automatyzacji może być np. cały zakład przemysłowy, sieć transportowa czy struktura organizacyjna. W węższym sensie, jakim będziemy się posługiwać dalej, obiektami automatyzacji są przede wszystkim „elementarne” **obiekty sterowane** tzn. obiekty, na które można wywierać sterowanie, przy czym obiekty te podlegają dodatkowo zakłóceniom. Obiekty sterowane o różnej naturze fizycznej będziemy rozważać mając na uwadze przede wszystkim cel — tj. sterowanie pewnych wyróżnionych sygnałów wyjściowych (wielkości sterowanych) tych obiektów.

·2.2· Co to jest model obiektu?

Mówiąc o modelu obiektu, najczęściej myślimy o sposobie przedstawienia obiektu. Może to być **model fizyczny**, np. makietą samolotu do badań w tunelu aerodynamicznym. Częściej jednak model ma inną naturę fizyczną niż obiekt oryginalny. Jeśli bowiem istnieją pewne analogie opisu różnych układów fizycznych, to można jeden układ modelować za pomocą innego, łatwiejszego i tańszego do realizacji lub wygodniejszego do badań. Tak zwany **model analogowy** (z reguły elektryczny) jest to układ zastępczy, w którym zachodzą zjawiska analogiczne (z punktu widzenia opisu matematycznego) do zjawisk w obiekcie oryginalnym. Modele analogowe należą do szerszej klasy **modeli symulacyjnych**. Przez symulację lub modelowanie rozumie się analizę właściwości obiektu na podstawie jego modelu. Jeśli badamy model analogowy — mówimy o symulacji lub modelowaniu analogowym. Dalszym rozszerzeniem tej idei jest stworzenie modelu o zupełnie innej naturze przez sformułowanie programu obliczeniowego dla komputera cyfrowego. Zjawiska zachodzące w komputerze nie mają już wtedy nic wspólnego z opisem obiektu oryginalnego, jednak przy poprawnym sformułowaniu programu otrzymamy dla odpowiedniego zbioru danych wejściowych zbiór wyników, reprezentują-

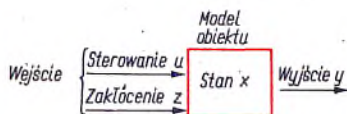
cy zachowanie się obiektu w określonych warunkach. Będzie to wtedy **symulacja cyfrowa** obiektu oryginalnego. W szczególności w przypadku symulacji cyfrowa może mieć postać symulacji modelu analogowego, a raczej jego opisu matematycznego. Dochodzimy w ten sposób do najbardziej ogólnego pojęcia **modelu matematycznego**. Terminem tym będziemy określać pełny opis matematyczny obiektu, np. układ równań różniczkowych opisujących działanie obiektu. Mówiąc „model” rozumiemy, że oznacza to w istocie tylko pewne wyobrażenie „prawdziwego” obiektu, na ogół nie obejmujące absolutnie wszystkich zjawisk zachodzących w obiekcie. Poprawność modelu matematycznego obiektu sterowanego wpływa w bardzo istotny sposób na jakość dalszych prac nad automatyzacją tego obiektu.

•2.3. Jakie najważniejsze wielkości występują w sformułowaniu modelu matematycznego obiektu?

Do sformułowania modelu matematycznego stosuje się wielkości dwóch rodzajów: zmienne i parametry. Zmienne są to te wielkości, które zawierają informację o zachowaniu się obiektu (stąd często mówi się o **sygnalach** związanych z obiektem). Zmienne z zasady są funkcjami czasu, tzn. sygnały reprezentujące te zmienne są pewnymi **przebiegami** czasowymi. Używa się też określenia **współrzedne** (obektu) dla podkreślenia, że wielkości te określają warunki, w jakich obiekt się znajduje w danej chwili. W odróżnieniu od zmiennych parametry charakteryzują pewne cechy i właściwości obiektu. Są to najczęściej wielkości stałe — np. współczynniki w równaniach opisujących obiekt. Parametry modelu wynikają wprost z parametrów fizycznych obiektu — stałych materiałowych, wymiarów geometrycznych itp. Parametry są to więc zwykle określone liczby (często mianowane), jednak w bardziej skomplikowanych przypadkach zachodzi potrzeba traktowania również parametrów jako funkcji czasu lub funkcji współrzędnych obiektu. Z powyższych określeń wynika także, że podstawową zmienną niezależną w opisie obiektów jest czas, względem którego odnosi się poszczególne wielkości.

•2.4. Na jakie grupy dzieli się zmienne modelu obiektu?

Rozróżnia się trzy podstawowe grupy zmiennych:
 — zmienne wejściowe (w tym zmienne sterujące i zmienne zakłócające);
 — zmienne wyjściowe;
 — zmienne stanu obiektu.



•2.4. Model obiektu i jego współrzedne

Zmienne sterujące (lub ogólnie wejściowe) oznacza się najczęściej literą u , zakłócające — z , zmienne wyjściowe — y , zaś zmienne stanu — x . Konfigurację modelu obiektu i jego współrzędnych (zmiennych) można przedstawić tak jak na rys. 2.4.

·2.5· Jaka jest rola zmiennych wejściowych i wyjściowych?

Zmienne wejściowe są to te zmienne, za których pośrednictwem można oddziaływać na obiekt, natomiast zmienne wyjściowe reprezentują sygnały stanowiące jedyne źródło informacji o danym obiekcie. Zmienne wejściowe reprezentują sterowanie, jeśli jest to oddziaływanie celowe, lub zakłócenie, jeśli jest to czynnik niepożądany; w niektórych przypadkach to rozróżnienie nie jest potrzebne.

Punkty, poprzez które można wprowadzać oddziaływania na obiekt, nazywa się zwykle wejściami tego obiektu, zaś punkty, w których można obserwować (mierzyć) zachowanie się obiektu — jego wyjściami. Zwracamy uwagę, że pojęcia „wejścia” i „wyjścia” nie muszą oznaczać „wpływanía” lub „wypływanía” strumieni energii czy substancji. Jeśli np. rozważanym obiektem jest zbiornik z cieczą, to jego wejściem (w sensie sterowania) może być odpływ, zaś wyjściem — np. poziom cieczy.

·2.6· Ile zmiennych wejściowych i wyjściowych może mieć układ?

Liczba zmiennych w układzie nie jest w zasadzie niczym ograniczona. Jest jednakże oczywiste, że racjonalny sposób opisu układu dąży do angażowania możliwie małej liczby istotnych zmiennych. W najprostszym przypadku układ może mieć jedną zmienną wejściową i jedną wyjściową — nazywa się to przypadkiem **skalarnym**. Często jednak cel, któremu ma służyć dany układ, wymaga wprowadzenia większej liczby zmiennych wejściowych (np. uwzględnienia oddzielnie sterowania i zakłócenia) lub zmiennych wyjściowych. W takich przypadkach mówi się o układzie o wielu wejściach i wielu wyjściach lub w skrócie — o układzie **wielowymiarowym**. W odróżnieniu od przypadku skalarnego określa się **wektory** zmiennych wejściowych lub wyjściowych, np.

$$\mathbf{u} = [u_1 \dots u_r]$$

$$\mathbf{y} = [y_1 \dots y_m]$$

rozumiejąc, że układ ma r wejść ($u_1, \dots, u_r$) i m wyjść ($y_1, \dots, y_m$). Oczywiście przypadek skalarny jest przypadkiem szczególnym przy $r = m = 1$.

·2.7· Jaki sens ma pojęcie stanu układu?

Jest to jedno z najważniejszych pojęć w teorii układów dynamicznych i teorii sterowania. Pojęcie to należy jednak rozumieć w bardziej ścisły sposób niż w języku potocznym. Pytając o stan danego układu (obektu), żądamy ścisłej informacji ilościowej, umożliwiającej określenie, co dzieje się z układem w danej chwili (nawet jeśli nie jest on poddany żadnym oddziaływaniom zewnętrznym) i jak może się on zachowywać w najbliższej przyszłości. Stan charakteryzuje układ odznaczający się pewnego rodzaju „pamięcią”, tzn.

stan zawiera informację zakumulowaną z całej przeszłości układu aż do danej chwili oraz nie może ulegać nagłym, skokowym zmianom.

Pojęcie stanu najlepiej zinterpretować na prostym przykładzie. Przypuśćmy, że mamy zbiornik z cieczą. Zawartość zbiornika (objętość cieczy) mówi nam wszystko o stanie zbiornika, a więc jest to zmienna stanu tego zbiornika. Wiedząc bowiem, ile jest w danej chwili cieczy w zbiorniku, nie musimy już pamiętać o tym, w jaki sposób i kiedy ta ciecz wpłynęła do zbiornika — zmienna stanu „pamięta” całą historię dopływów i odpływów ze zbiornika i bilansuje tę historię. Dopływ lub odpływ można uznać za zmienną sterującą zbiornika; łatwo spostrzec, że ustalając np. określony dopływ, mierzony w jednostkach objętości na jednostkę czasu, regulujemy stan, ale nie możemy go zmieniać nagle — co najwyżej wpływamy na prędkość zmian stanu. Zauważmy również, że znajomość stanu aktualnego jest niezbędna, jeśli interesujemy się przyszłym zachowaniem układu — np. jeśli chcemy określić, w jaki sposób zbiornik opróżni się z cieczy przy swobodnym jej wypływie. Stan zbiornika jest jego zmienną wewnętrzną, bowiem zawartość zbiornika można określić tylko pośrednio (jeśli nie wypuścimy całej cieczy i nie zmierzymy jej objętości) — np. mierząc poziom cieczy w zbiorniku lub ciśnienie na jego dnie. Poziom lub ciśnienie można oczywiście uznać za zmienne wyjściowe; nie zawsze są one proporcjonalne do stanu (w naszym przykładzie zależą od kształtu zbiornika).

Przykład zbiornika uwidacznia kilka istotnych spraw. Najważniejszą jest spostrzeżenie, że wybór zmiennej stanu jest w gruncie rzeczy arbitralny. Zawartość zbiornika mówi nam wszystko o jego stanie pod warunkiem, że nas właśnie interesuje tylko zawartość, nie zaś np. zjawiska hydrodynamiczne w zbiorniku (fale, zawirowania, ich rozkłady w przestrzeni itd.), lub temperatura lub skład cieczy (zwłaszcza jeśli w zbiorniku następuje mieszanie kilku różnych cieczy). Tak więc, określając zmienne stanu jakiegoś układu, trzeba wiedzieć, jak będzie on wykorzystywany i na co w związku z tym należy zwracać uwagę. Drugie spostrzeżenie to fakt, że rozważany zbiornik może się przelać lub całkowicie opróżnić. Oznacza to, że stan może (choć nie musi) podlegać pewnym ograniczeniom fizycznym. Zauważmy wreszcie, że znajomość stanu daje nam bardzo wiele, ale jeszcze więcej wiedzielibyśmy o układzie, gdybyśmy znali związki zmiennej stanu z innymi ważnymi zmiennymi. Można więc przypuszczać, że w opisie układu (w jego modelu matematycznym) kluczową rolę będzie odgrywał związek (np. równanie) rządzący zachowaniem się zmiennej stanu. W przykładzie zbiornika można taki związek podać bardzo łatwo w postaci równania różniczkowego.

$$\frac{dx(t)}{dt} = u(t)$$

w którym $x(t)$ oznacza wartość zmiennej stanu w danej chwili (objętości cieczy w litrach), zaś $u(t)$ oznacza wypadkowy dopływ w danej chwili (w litrach na sekundę). Równie łatwo można napisać rozwiązanie powyższego równania (zmienna τ jest zmienną całkowania)

$$x(t) = \int_{-\infty}^t u(\tau) d\tau$$

lub

$$x(t) = x(0) + \int_0^t u(\tau) d\tau$$

Pierwsze z tych wyrażeń pokazuje, że zawartość zbiornika jest całką wypadkowego dopływu, liczoną za cały przeszły okres czasu (w czasie „ $-\infty$ ” zbiornik był pusty). Drugie wyrażenie jest bardziej dogodne i podkreśla postulowaną przez nas właściwość zmiennej stanu: jeśli mianowicie rozpatrujemy, co dzieje się w zbiorniku od chwili 0, to przede wszystkim musimy znać stan $x(0)$ — zwalnia to nas z potrzeby interesowania się przeszłością; dalsze zachowanie natomiast określa całka wielkości $u(t)$ liczona od chwili początkowej 0 do bieżącej t , przy czym dla $u = 0$, i $t > 0$, stan $x(0)$ pozostałby stały („zapamiętany”) również w następnych chwilach czasu t .

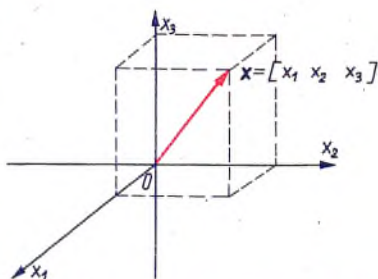
Na koniec podkreślmy, że zupełnie analogiczne właściwości do omówionego zbiornika miałyby takie układy jak kondensator ładowany prądem elektrycznym (zmienną stanu jest ładunek zgromadzony w kondensatorze), kalorymetr bilansujący ciepło doprowadzone i odprowadzone, czy wreszcie magazyn towaru lub zasób finansowy. Wspólną cechą wymienionych układów jest fakt magazynowania (zachowywania, pamiętania) pewnej wielkości fizycznej — masy, ładunku, energii, ciepła, kapitału.

•2.8• Czy stan układu opisuje się tylko jedną zmienną?

Przedstawiony przykład zbiornika jest dość rzadkim przypadkiem, kiedy wystarcza pojedyncza zmienna dla określenia stanu układu. Zazwyczaj trzeba określić większą liczbę zmiennych stanu. Ponieważ zmienne stanu są związane z istnieniem elementów magazynujących np. energię, więc liczba zmiennych stanu jest równa liczbie **niezależnych** elementów magazynujących.

•2.9• Co to jest wektor stanu i przestrzeń stanów?

Pojęcie wektora stanu jest uogólnieniem w stosunku do pojedynczej zmiennej stanu. Jeśli układ jest opisany tylko jedną zmienną stanu, to jej wartości są reprezentowane przez liczby skalarnie (rzeczywiste). W przypadku większej liczby zmiennych stanu nie można określić konkretnego stanu za pomocą jednej liczby, lecz za pomocą zbioru liczb repre-



·2.9· Interpretacja przestrzeni stanów i wektora stanu w przypadku $n = 3$

zentuujących wartości poszczególnych zmiennych. Można to interpretować w taki sposób, że stan ma sens **wektora**, określonego w **przestrzeni stanów**, n -wymiarowej, jeśli istnieje n zmiennych stanu. Ośiami (współzrzednymi) przestrzeni stanów są więc poszczególne współzrzedne (zmienne) stanu. Każdy punkt przestrzeni stanów reprezentuje określony stan rozumiany jako zbiór wartości wszystkich zmiennych stanu układu. Można więc zapisać symbolicznie wektor stanu układu o n zmiennych stanu $x_1, x_2, \dots, x_n$ jako

$$\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]$$

Na rysunku 2.9 podano interpretację przestrzeni stanów i wektora stanu w przypadku $n = 3$.

·2.10· Co to jest układ statyczny?

Jest to układ, w którym nie można określić ani jednej zmiennej stanu. W układzie statycznym nie ma elementów magazynujących, lecz są tylko elementy rozpraszające energię. Przykładami układów statycznych są np. sieć elementów rezystancyjnych w elektrotechnice lub układ dźwigni mechanicznych (o pomijalnej bezwładności) w mechanice. Układy takie nazywa się też **bezinercyjnymi**. Niektóre obiekty sterowane można przedstawić jako układy statyczne. Do opisanania takich obiektów potrzebna jest tylko zależność funkcyjna pomiędzy wejściami a wyjściami, obowiązująca zwykle w sposób jednoznaczny w każdej chwili. Zależność zmiennej wyjściowej od zmiennej wejściowej w układzie statycznym nazywa się **charakterystyką statyczną**.

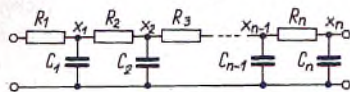
·2.11· Co to jest układ dynamiczny?

W szerokim sensie układem dynamicznym nazywa się każdy układ, w którym rozpatruje się przebiegi zmiennych jako funkcji czasu. Ścisłej, jako układ dynamiczny rozumie się układ, w którym można określić przynajmniej jedną zmienną stanu. Jest to rozróżnienie zasadnicze, gdyż w układzie statycznym, nie posiadającym zmiennych stanu, rozpatrywanie przebiegów czasowych nie mówi nic o istocie właściwości układu, skoro te właściwości opisuje wyczerpująco charakterystyka statyczna. W układzie dynamicznym w ścisłym sensie wchodzi w grę zależności pomiędzy przebiegami czasowymi zmiennych, a nie ich wartościami chwilowymi. W szczególności układ dynamiczny bez wymuszeń (tzn. przy zerowych sygnałach na wejściach) także wykazuje określony przebieg zmiennych stanu, zależny od wartości początkowych tych zmiennych i od właściwości układu.

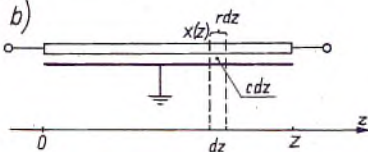
·2.12· Co to jest rząd układu dynamicznego?

Jest to liczba zmiennych stanu, czyli wymiar wektora stanu układu. Rząd będzie oznaczany przez n . Układ statyczny ma rząd równy zeru, układ dynamiczny ma rząd $n \geq 1$.

a)



b)



•2.14• Układ o parametrach skupionych (a) i o parametrach rozłożonych (b)

•2.13• Czy rząd układu może być nieskończony?

Tak; układ dynamiczny, dla którego $n \rightarrow \infty$, nazywa się układem o **parametrach rozłożonych**, w odróżnieniu od układu o **parametrach skupionych**, w którym n może być bardzo duże, lecz skończone.

•2.14• Jaka jest różnica pomiędzy układami o parametrach skupionych i rozłożonych?

Układ o parametrach skupionych ma skończoną liczbę elementów magazynujących. Elementy te można kolejno ponumerować, a więc i zlokalizować. Natomiast „rozłożoność” parametrów oznacza, że nie można w układzie wyróżnić odrębnych elementów, gdyż są one usytuowane nieskończenie „gęsto” względem siebie. Obrazuje to rys. 2.14: w układzie (a) można odróżnić poszczególne człony RC połączone ze sobą, natomiast układ (b) przedstawia tzw. kontinuum RC, w którym można określić odpowiednie parametry tylko przez podanie wartości rezystancji i pojemności na jednostkę długości z : $r = \frac{dR}{dz} \left[\frac{\Omega}{m} \right]$, $c = \frac{dC}{dz} \left[\frac{F}{m} \right]$.

Jeśli nawet sumaryczna długość jest skończona (równa Z), to i tak układ wcale nie będzie się zachowywał tak jak pojedynczy człon o parametrach wypadkowych równych $R = rZ$ i $C = cZ$, ponieważ ma on postać nieskończonego szeregu elementarnych członów o parametrach $r dz$ i $c dz$, rozłożonych nieskończenie gęsto. Dla wyczerpującego opisu układu nie wystarczy więc podanie przebiegów napięć na n kondensatorach, tak jak w przypadku (a): $x_1(t), \dots, x_n(t)$, lecz trzeba podać rozkłady napięcia $x(z)$ wzdłuż długości z , dla poszczególnych chwil t , tzn. podać funkcję $x(z, t)$. Efekt rozłożoności parametrów znacznie wyraźniej występuje w układach z wymianą ciepła lub przepływem płynów, przy czym w grę mogą wchodzić rozkłady względem trzech współrzędnych przestrzennych, a nie tylko względem długości.

Należy zaznaczyć, że w istocie każdy układ fizyczny jest układem o parametrach rozłożonych. Przyjęcie modelu o parametrach skupionych jest więc tylko przybliżeniem.

•2.15• Czy zmienne stanu układu mogą przybierać wartości dowolne?

Jak już wspomniano, często wartości zmiennych stanu są ograniczone do pewnych przedziałów osi liczbowej. Znacznie istotniejsze jest jednak stwierdzenie, czy zmienne te mogą przybierać wszystkie możliwe wartości ze swojego przedziału ograniczoności, czy też tylko wybrane, określone wartości. W pierwszym przypadku mówi się o **zmiennych ciągłych**, w drugim — o **zmiennych dyskretnych**. Często można spotkać określenia **sygnały analogowe** lub odpowiednio **cyfrowe**.

•2.16• Czy zmienna niezależna — czas — może podlegać dyskretyzacji?

Tak, również czas może być w układzie dynamicznym zdykretyzowany. Oznacza to, że wartości sygnałów w układzie zmieniają się tylko w wyróżnionych chwilach lub że dla opisu właściwości układu wystarcza znajomość wartości sygnałów w wyróżnionych chwilach czasu. Układy, w których czas jest zmienną ciągłą, nazywa się **układami ciągłymi**, natomiast układy, w których czas rozważa się jako zmienną dyskretną, nazywa się **układami dyskretnymi** lub **impulsowymi**. Najczęściej w układach impulsowych operacja impulsowania (inaczej próbkowania), polegająca na wyznaczaniu wyróżnionych chwil czasu, ma charakter okresowy. W takiej sytuacji przebiegi czasowe sygnałów można przedstawiać w postaci ciągów wartości, np. $x(kT_p)$ (przy czym $k = 0, 1, 2, \dots$, zaś T_p jest okresem impulsowania), zamiast w postaci funkcji argumentu ciągłego, np. $x(t)$.

•2.17. Czy parametry układu mogą zależeć od czasu?

Tak, chociaż bowiem uznaliśmy początkowo, że parametry układu są to na ogół stałe, to jednak w rzeczywistości cechy układu określone jego parametrami mogą podlegać zmianom w czasie. W przypadku gdy zmienność parametrów w czasie nie występuje lub jest do pominięcia — mamy do czynienia z układami stacjonarnymi, w przypadku układów o parametrach zmiennych będziemy mówić o układach niestacjonarnych.

•2.18. Jakie układy nazywa się liniowymi?

Układy liniowe są to takie układy, których opis ma postać zależności liniowych. Postulat liniowości wymaga, aby była spełniona zasada superpozycji. Oznacza to, że reakcja układu liniowego na wymuszenie o postaci $u = c_1 u_1 + c_2 u_2$ ma postać $y = c_1 y_1 + c_2 y_2$, przy czym y_1 i y_2 stanowią wynik oddziaływania oddzielnych wymuszeń u_1 i u_2 , a c_1 i c_2 są stałymi dowolnymi. Opis (np. równania) układu liniowego można dość łatwo rozpoznać, ponieważ w opisie takim nie mogą występować żadne operacje nieliniowe na zmiennych układu (a więc np. iloczyny lub potęgi zmiennych), a parametry układu (współczynniki równań) nie mogą zależeć od zmiennych.

•2.19. Czy układy liniowe często spotyka się w praktyce?

W rzeczywistości układy liniowe nie istnieją. Postulat liniowości jest bowiem niezwykle ostrym warunkiem, w szczególności wymaga, aby żadna ze zmiennych układu nie podlegała żadnym ograniczeniom. Stoi to w oczywistej sprzeczności z doświadczeniem, które uczy, że w otaczającym nas świecie nie istnieją zmienne fizyczne, które mogłyby przyjmować wartości dowolnie duże. Z drugiej strony, przy badaniu zjawisk w świecie mikroskopowym okazuje się z zasady, że występuje pewna kwantyzacja, „ziarnistość” (energii, materii, ładunku elektrycznego), co również nie mieści się w ramach warunków liniowości.

Wobec powyższych uwag należy stwierdzić, że można mówić tylko o modelu liniowym układu — i to wyłącznie o modelu abstrakcyjnym, matematycznym. Na to zaś, aby taki model można było zastosować do danego układu fizycznego, należy znaleźć przynajmniej pewien obszar wartości zmiennych, w którym postulat liniowości jest spełniony z zadowalającą dokładnością.

·2.20· Czy modele liniowe są często stosowane?

Odpowiedź na to pytanie jest twierdząca, pomimo zastrzeżeń wymienionych w pytaniu poprzednim. Spełnienie przez model postulatu liniowości daje ogromne korzyści, jeśli chodzi o łatwość analizy i obliczeń. Dla układów (a raczej modeli) liniowych istnieją ściśle sformułowania wszystkich najważniejszych właściwości i ściśle rozwiązania analityczne dotyczące większości przypadków praktycznych. Dlatego też modele liniowe stosuje się powszechnie do analizy różnych układów, przy czym najczęściej pomija się nawet zastrzeżenia co do zakresu ważności postulatu liniowości, jeśli ten zakres liniowości pokrywa się z typowym, użytecznym, „rozsądnym” zakresem pracy danego układu. Co więcej, nazywa się (choć nie zbyt ściśle) „układami liniowymi” układy, dla których istnieje przynajmniej pewien zakres pracy „liniowej”.

·2.21· Jaki jest najprostszy układ statyczny?

Najprostszym przykładem układu statycznego jest układ, którego model można przedstawić w postaci zależności

$$y = ku$$

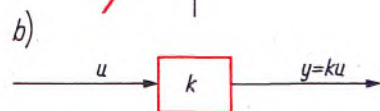
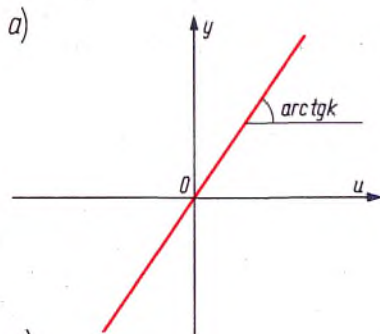
przy czym k jest stałym współczynnikiem. Jest to oczywiście zależność liniowa, a człon opisany taką zależnością nazywa się **członem proporcjonalnym**. Człon ten realizuje mnożenie przez stałą, a więc powoduje tylko zmianę skali zmiennej wyjściowej y w stosunku do wejścia u . Współczynnik k może być mianowany lub nie. W pierwszym przypadku człon ten często nazywa się **przetwornikiem**, np. jeśli u jest temperaturą, a y napięciem — mamy do czynienia z przetwornikiem termoelektrycznym, zaś współczynnik k ma wymiar $[V/K]$. W sytuacji, gdy obydwie zmienne u i y mają ten sam wymiar fizyczny, człon proporcjonalny nazywa się często **wzmocniaczem** (zwłaszcza przy $|k| > 1$) lub **dzielnikiem** albo **łumiakiem** (przy $|k| < 1$). Charakterystyka statyczna członu proporcjonalnego ma oczywistą postać linii prostej o współczynniku kątowym k (rys. 2.21).

·2.22· Co to jest węzeł sumujący?

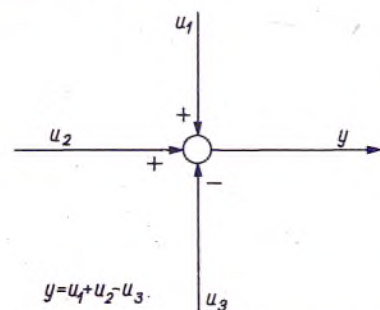
W przypadku wielu wejść — z członu proporcjonalnego otrzymuje się **człon sumujący** opisany równaniem

$$y = k_1 u_1 + k_2 u_2 + \dots + k_r u_r$$

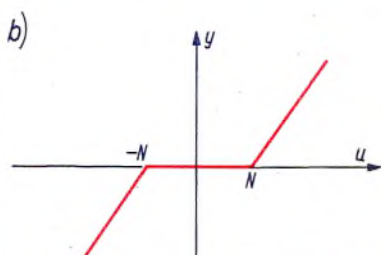
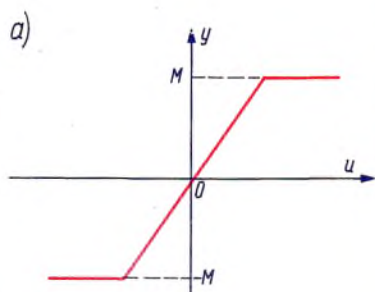
Szczególnie przydatne jest wyróżnienie sytuacji, gdy współczynniki k_i ($i = 1, \dots, r$) są równe tylko $+1$ lub -1 . Mówi



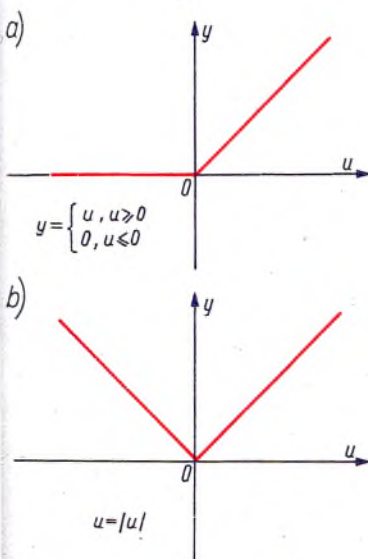
·2.21· Charakterystyka statyczna członu proporcjonalnego i jego symbol graficzny



·2.22· Przykład oznaczenia węzła sumującego



•2.23A• Elementarne typy nieliniowości: a) ograniczenie; b) strefa nieczułości



•2.23B• Elementarne typy nieliniowości o dwu zakresach pracy: a) typu diody idealnej; b) wartość bezwzględna

się wtedy o węzle sumującym (lub różnicowym) i oznacza się go symbolicznie w sposób pokazany na rys. 2.22.

•2.23• Jakie są elementarne typy nieliniowości układów statycznych?

Za najbardziej elementarne typy nieliniowości należy uznać przypadki z jedną zmienną wejściową u i wyjściową y , w których nieliniowość polega na istnieniu kilku zakresów liniowości na charakterystyce statycznej $y(u)$.

Na rysunku 2.23A pokazano dwa najbardziej elementarne przypadki: tzw. efekt **nasycenia** lub **ograniczenia** na poziomie $\pm M$ oraz efekt **strefy nieczułości** o szerokości $\pm N$. Zauważmy, że są to w ogóle najbardziej elementarne pojęciowo efekty nieliniowe (ograniczonność dużych wartości y , nierozróżnialność małych wartości u). Do grupy nieliniowości elementarnych można również zaliczyć charakterystyki typu **diody idealnej** lub **wartości bezwzględnej** (rys. 2.23B), na których istnieją tylko po dwa zakresy ($u > 0$ i $u < 0$) o różnych właściwościach.

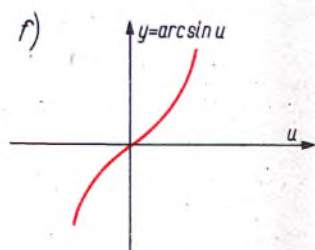
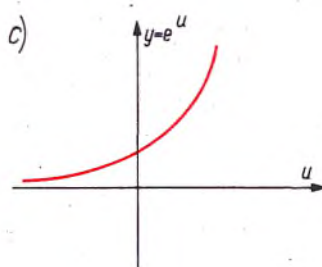
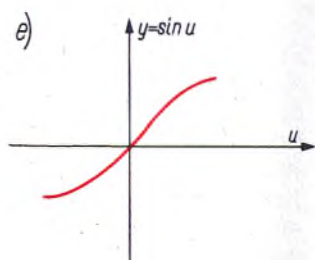
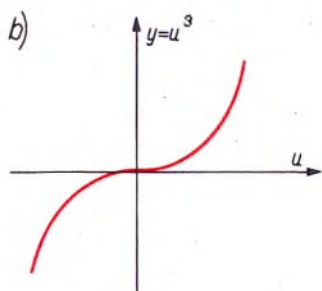
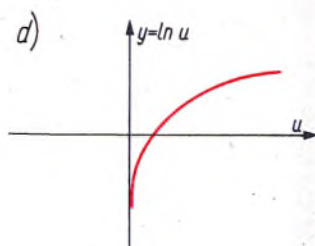
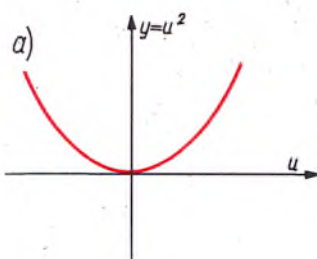
Pokazane charakterystyki są oczywiście wyidealizowane, ponieważ w praktyce rzadko spotyka się sytuację, w których występują tak wyraźne różnice właściwości w poszczególnych zakresach charakterystyki statycznej, a więc tak wyraźne „zagięcia” wykresu $y(u)$.

•2.24• Czy do opisu członów nieliniowych zawsze jest potrzebna charakterystyka statyczna w postaci wykreślnej?

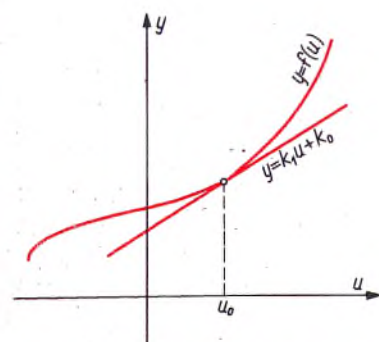
Nie, chociaż wykreślne przedstawienie charakterystyki statycznej jest zawsze bardzo pogładowe, a często jest to najprostszy sposób opisu właściwości członów nieliniowych (np. opis inny niż graficzny członów typu „strefa nieczułości” czy „dioda” staje się kłopotliwy i mało przejrzysty). Trzeba jednak podkreślić, że jeśli to jest możliwe, należy starać się podać opis analityczny nieliniowości. Opis analityczny jest szczególnie istotny przy charakterystykach nieliniowości „gładkich”, gdzie, przeciwnie, wykres może dostarczać niewiele informacji o opisywanym członie w porównaniu z wyrażeniem analitycznym. Za najbardziej typowe nieliniowości „gładkie” można uznać np. $y = u^2$, $y = u^3$, $y = e^u$, $y = \ln u$, $y = \sin u$, $y = \arcsin u$ — i podobne, opisywane dobrze znanymi funkcjami. Na rysunku 2.24 pokazano kształt charakterystyk wymienionych nieliniowości.

•2.25• Co to jest linearyzacja charakterystyk nieliniowych?

Linearyzacja jest to uproszczenie modelu nieliniowego w taki sposób, że charakterystykę nieliniową przybliża się lokalnie, tzn. w pewnym obszarze, odpowiednio dobraną zależnością liniową. Cel takiego zabiegu jest oczywisty — znaczne uproszczenie opisu i analizy. Zasadę linearyzacji można zilustrować graficznie (rys. 2.25). Jeśli nieliniowość opisuje się charakterystyką statyczną $y = f(u)$, to w pewnym punkcie tej charakterystyki, np. przy $u = u_0$, można



•2.24• Kształt charakterystyk statycznych typowych nieliniowości „gładkich”



•2.25• Ilustracja zasady linearyzacji

poprowadzić styczną do charakterystyki i uznać, że w pobliżu tego punktu charakterystyka pokrywa się ze styczną, a więc ma postać $y = k_1 u + k_0$, przy czym współczynnik k_1 odpowiada nachyleniu stycznej, zaś k_0 jest przesunięciem względem początku układu współrzędnych. Ten sam efekt można przedstawić analitycznie, ponieważ funkcję $y = f(u)$ można rozwinąć w szereg potęgowy (szereg Taylora)

$$y(u) = f(u_0) + \left. \frac{df(u)}{du} \right|_{u=u_0} (u-u_0) + \frac{1}{2} \left. \frac{d^2f(u)}{du^2} \right|_{u=u_0} (u-u_0)^2 + \dots$$

Jeśli w szeregu nieskończonym pominąć wszystkie wyrazy poza dwoma pierwszymi — otrzymuje się przybliżenie liniowe funkcji $y = f(u)$. Na to, aby można było przyjąć to przybliżenie, powinny być spełnione pewne warunki: odchylenie $(u-u_0)$ punktu „bieżącego” u od punktu u_0 , wokół którego dokonujemy rozwinięcia, musi być dostatecznie małe, tak aby człony $(u-u_0)^2$, $(u-u_0)^3$ itd., pojawiające się w dalszych wyrazach rozwinięcia, były pomijalnie małe wobec

wyrazów pierwszych. Tak więc linearyzacja jest możliwa tylko lokalnie, w otoczeniu wybranego punktu u_0 (np. „punktu pracy” układu) i ma sens tylko dla **małych odchy-
leń** od tego punktu. Ponadto, aby przybliżenie liniowe miało sens — powinna istnieć pochodna $\frac{df(u)}{du}$ w punkcie u_0 .

Ten drugi warunek oznacza „gładkość” charakterystyki $y = f(u)$ w punkcie u_0 , tzn. brak skoków lub załamania charakterystyki statycznej w tym punkcie lub inaczej mówiąc istnienie w tym punkcie stycznej do charakterystyki. Jeśli te warunki są spełnione, to rzeczywiście można napisać

$$y(u) \approx f(u_0) + \left. \frac{df(u)}{du} \right|_{u=u_0} (u - u_0) = k_1 u + k_0$$

przy czym łatwo sprawdzić, że $k_1 = \left. \frac{df(u)}{du} \right|_{u=u_0}$

oraz $k_0 = f(u_0) - k_1 u_0$.

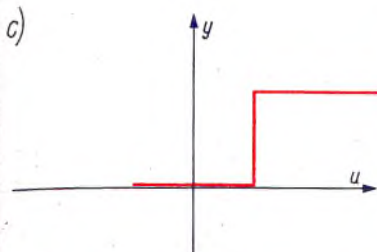
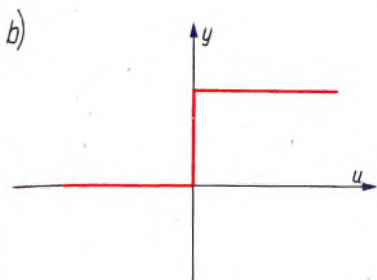
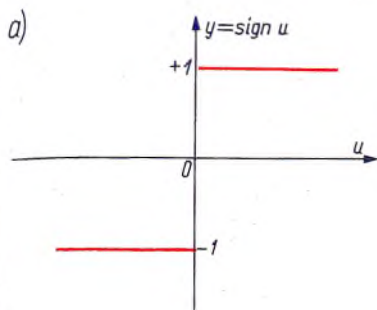
Powyższa metoda linearyzacji dla małych odchylenia jest często stosowana — również w przypadku zależności od kilku zmiennych: $y = f(u_1, u_2, \dots, u_r)$.

•2.26• Kiedy nie można stosować linearyzacji?

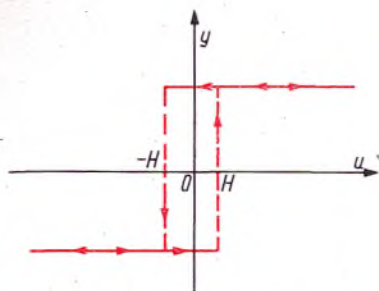
Argumentem przesądzającym o dopuszczalności linearyzacji jest żądana dokładność przybliżenia liniowego lub — co równoważne — żądany przedział zmienności, w którym linearyzacja ma obowiązywać. Jeśli postawić żądanie bardzo dużej dokładności przybliżenia, to może się okazać, że dokładność taką otrzymuje się tylko w bardzo wąskim przedziale u . Z drugiej strony — jeśli zażądać, aby przybliżenie liniowe było stosowane w szerokim zakresie zmienności u , to otrzymuje się małą dokładność, czyli dużą rozbieżność w stosunku do charakterystyki pierwotnej. Tak więc kwestię dopuszczalności linearyzacji trzeba rozstrzygnąć w konkretnej sytuacji. W niektórych przypadkach można z góry powiedzieć, że linearyzacja nie ma sensu. Zachodzi to zwłaszcza wtedy, gdy w normalnych warunkach eksploatacji pracuje się głęboko w strefie nasycenia charakterystyki statycznej, co jest typowe np. dla elementów typu przekaźnikowego tj. elementów, których charakterystyki statyczne wykazują skoki (nieciągłości).

•2.27• Co to jest element przekaźnikowy?

Jest to element, którego charakterystyka statyczna wykazuje nieciągłości (skoki) wartości y . Właściwości przekaźnikowe wykazują wiele elementów; najprostszym przykładem są przekaźniki elektromagnetyczne zmieniające stan skokowo — zwarcie lub rozwarcie styków. Prócz tego można wymienić elementy czysto mechaniczne (sprzęgła, zapadki) lub elektroniczne (tzw. układy przełączające lub logiczne) charakteryzujące się skokowymi zmianami sygnału wyjściowego.



•2.27A• Charakterystyki przekaźnikowe: a) charakterystyka idealna; b) charakterystyka niesymetryczna; c) charakterystyka przesunięta



•2.27B• Charakterystyka przekąźnikowa z histerezą (przebieg idealny)

Za najbardziej elementarną, „idealną” charakterystykę przekąźnikową można uważać charakterystykę typu $y = \text{sign } u$, pokazaną na rys. 2.27A (a). Funkcja „sign” jest równa $+1$ lub -1 , zależnie od znaku u , zaś przy $u = 0$ nie jest określona (można umownie przypisać jej w tym punkcie wartość 0). Najczęściej spotyka się pewne odmiany tej charakterystyki, pokazane na rys. 2.10b i 2.10c. Wyróżnionych wartości sygnału wyjściowego y może być więcej niż dwie — używa się więc określeń „element dwustawny” (dwpółożeniowy), „trójstawny” (trójpołożeniowy) itd.

Niemal zawsze elementy przekąźnikowe wykazują efekt tzw. **histerezy**, czyli różnicy wartości progów „zadziałania” i „zwolnienia”. Elementarną charakterystykę przekąźnikową z histerezą pokazano na rys. 2.27B. Jest to zjawisko bardzo ciekawe, ponieważ wprowadza nie tylko nieciągłość, ale co więcej — niejednoznaczność charakterystyki, zależnie od kierunku zmian wartości u , co na rysunku zaznaczono strzałkami.

•2.28. W jaki sposób można łączyć człony nieliniowe?

Człony nieliniowe z jednym wejściem i jednym wyjściem można łączyć trzema sposobami: **szeregowo**, **równolegle** i na zasadzie **sprzężenia zwrotnego**. Na rysunku 2.28A pokazano te trzy typy połączeń na przykładzie dwu członów opisanych zależnościami $y = f_1(u_1)$, $y_2 = f_2(u_2)$.

Przy połączeniu szeregowym obowiązuje równość $y_1 = u_2$, stąd zależność wypadkowa stanowi „złożenie” f_2 i f_1 :

$$y = y_2 = f_2(y_1) = f_2[f_1(u_1)] = f_{sz}(u_1)$$

W przypadku połączenia równoległego obowiązują warunki

$$u_1 = u_2 = u, \quad y = y_1 + y_2, \quad \text{stąd}$$

$$y = y_1 + y_2 = f_1(u) + f_2(u) = f_r(u)$$

W przypadku połączenia ze sprzężeniem zwrotnym zachodzą związki $u_1 = u + y_2$ oraz $u_2 = y_1 = y$. Wobec tego

$$y = f_1(u_1) = f_1[u + f_2(y)]$$

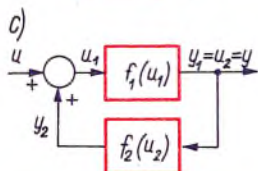
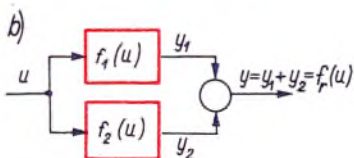
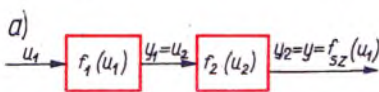
Jak widać, jest to przypadek zależności uwikłanej (y po obu stronach równości). Wynika to z „pętli” oddziaływań $u_1 \rightarrow y \rightarrow y_2 \rightarrow u_1 \dots$. Układy, w których występują tylko połączenia szeregowe lub równoległe (zależność ma tam postać jawną) nazywa się **układami otwartymi**, zaś układy z pętlami (o niejawnym, uwikłanym opisem) nazywa się układami **zamkniętymi** (lub układami z rekursją).

Próbując rozwikłać powyższą zależność można zapisać funkcję odwrotną do funkcji f_1 jako

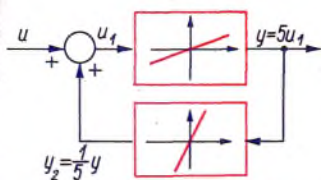
$$f_1^{-1}(y) = u + f_2(y)$$

Stąd otrzymamy

$$u = f_1^{-1}(y) - f_2(y) = g(y)$$



•2.28A• Połączenia członów: a) szeregowo; b) równoległe; c) na zasadzie sprzężenia zwrotnego



• 2.28B • Nierealny układ ze sprzężeniem zwrotnym

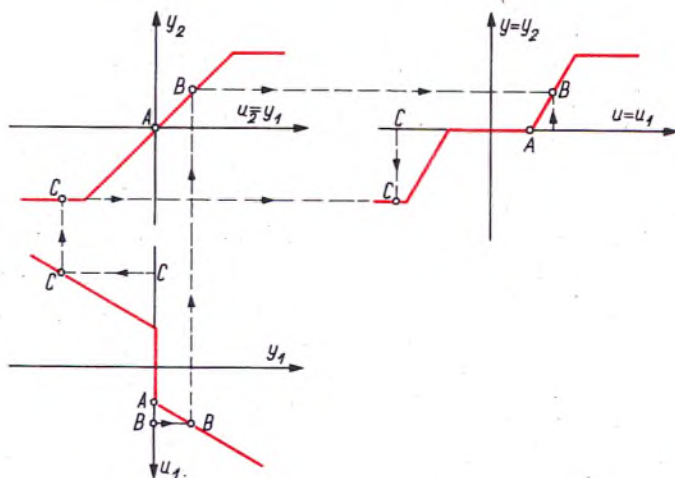
Pisząc ponownie funkcję odwrotną do g otrzymalibyśmy szukaną zależność jako

$$y = g^{-1}(u)$$

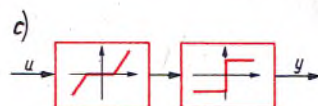
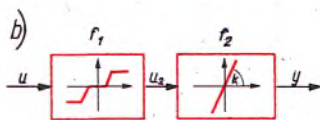
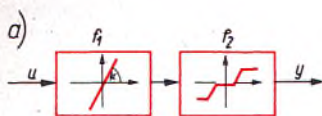
Powstaje jednak pytanie, czy powyższe przekształcenia są zawsze możliwe do zrealizowania. Okazuje się, że nie, ponieważ po pierwsze — funkcja odwrotna f_1^{-1} może nie istnieć (tzn. może nie być jednoznaczna), jeśli funkcja pierwotna nie jest różnowartościowa, po drugie zaś — może zajść sytuacja, w której funkcja g może być tożsamościowo równa zero. Przypadek pierwszy jest nieco mniej groźny, ponieważ niejednoznaczność zależności rozumianej nie jako funkcja, lecz jako charakterystyka statyczna — jest dość łatwa do uwzględnienia przy użyciu metod wykreślnych. Przypadek drugi jest znacznie gorszy; nie byłoby niczym niezwykłym, gdyby zaszła równość $f_1^{-1}(y) = f_2(y)$, a wtedy $g(y) = 0$ dla wszystkich y , czyli jedyny sensowny stan może mieć miejsce przy $u = 0$. Na rysunku 2.28B pokazano pozornie prosty układ złożony z dwu członów proporcjonalnych: $y = 5u_1$, $y_2 = \frac{1}{5}y$. Jak łatwo sprawdzić, w układzie tym obowiązuje zależność wypadkowa $y = 5(u + \frac{1}{5}y)$. Jednakże równość ta może być spełniona tylko przy $u = 0$ albo $y \rightarrow \infty$, a więc praktycznie układ taki nie ma sensu.

• 2.29 • Jakie można podać przykłady połączenia szeregowego członów nieliniowych?

W przypadku charakterystyk statycznych podanych wykreślnie można łatwo uzyskać charakterystykę wypadkową dwóch charakterystyk przez odpowiednie przerzutowanie punktów, tak aby oś y_1 (wyjście pierwszego członu) przerzutować na oś u_2 (wejście drugiego członu). Pokazano to na rys. 2.29A na przykładzie połączenia szeregowego członu ze strefą nieczułości z członem wprowadzającym nasycenie.



• 2.29A • Konstrukcja wypadkowej charakterystyki statycznej członów połączonych szeregowo



•2.29B• Przykłady połączeń szeregowych i odpowiednich charakterystyk wypadkowych: a) człon liniowy i człon ze strefą nieczułości i nasyceniem; b) jak wyżej, lecz w odwrotnej kolejności; c) człon ze strefą nieczułości i idealny element przekaźnikowy dwustawny, którego połączenia tworzą idealny element trójstawny

Jak widać, wypadkowa charakterystyka wykazuje obydwie cechy. Warto zwrócić uwagę, że w przypadku ogólnym kolejność łączenia poszczególnych członów ma istotne znaczenie — wynika to także z zapisu analitycznego, ponieważ ogólnie $f_2[f_1(u)] \neq f_1[f_2(u)]$. Wyjątkiem jest m.in. sytuacja, gdy dwie funkcje nieliniowe są odwrotne względem siebie, np. przy $f_1 = u^2$ i $f_2 = \sqrt{u_2}$ otrzymujemy $f_2[f_1(u_1)] = \sqrt{u_1^2} = u_1$, ale również $f_1[f_2(u_2)] = (\sqrt{u_2})^2 = u_2$ — a więc w obydwu przypadkach zależność wypadkowa jest liniowa. Z właściwości tej korzysta się np. przy linearyzacji charakterystyki przetworników pomiarowych.

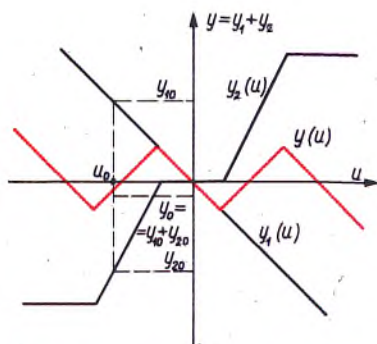
Na rysunku 2.29B pokazano kilka innych prostych przykładów połączeń szeregowych. Zauważmy, że jeśli jednym z członów połączenia szeregowego jest człon liniowy typu $y = ku$, to jest to równoważne tylko zmianie skali odpowiedniej osi.

Rozważanie połączenia szeregowego członów opisanych analitycznie jest mało ciekawe, ponieważ charakterystykę wypadkową w postaci analitycznej otrzymuje się natychmiast, np. przy $f_1 = u^2$, $f_2 = e^{u_2}$ uzyskuje się $y_2 = f_2[f_1(u_1)] = e^{u_1^2}$.

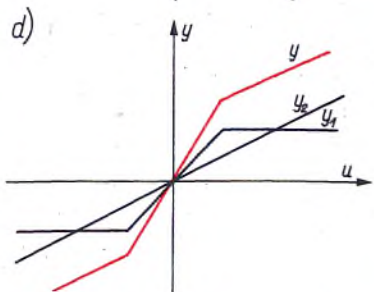
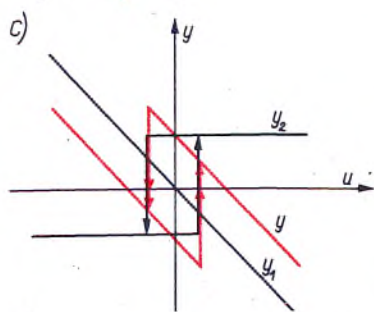
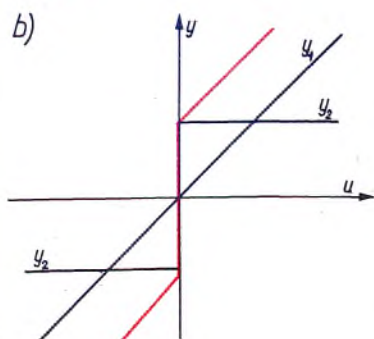
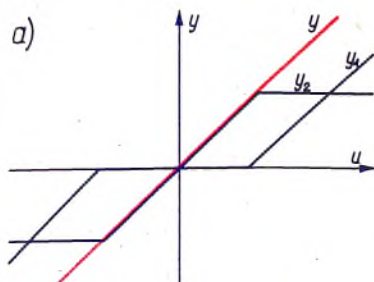
•2.30• Jakie można podać przykłady połączenia równoległego członów nieliniowych?

Podobnie jak poprzednio, konstrukcja graficzna, przy charakterystykach danych wykreślnie, jest bardzo prosta. Sumowanie wyjść poszczególnych członów przy wspólnym wejściu sprowadza się bowiem do sumowania rzędnych charakterystyk poszczególnych członów, narysowanych we wspólnym układzie współrzędnych.

Na rysunku 2.30A pokazano tę konstrukcję na przykładzie połączenia członu liniowego i członu ze strefą nieczułości i nasyceniem. Na rysunku 2.30B pokazano kilka innych prostych przykładów. Jak widać, w szczególnych przypadkach można „uzupełnić” jedną charakterystykę nieliniową inną charakterystyką, tak że w sumie otrzymuje się charakterystykę liniową.



•2.30A• Konstrukcja wypadkowej charakterystyki statycznej połączenia równoległego członu liniowego $y_1(u) = -ku$ i członu ze strefą nieczułości i nasyceniem $y_2(u)$



•2.30B• Przykłady charakterystyk połączeń równoległych ($y = y_1 + y_2$):
 a) „kompensacja” strefy nieczułości i nasycenia; b) połączenie członu liniowego i przekątnikowego idealnego; c) połączenie członu liniowego i przekątnikowego z histerezą; d) połączenie członu liniowego i nieliniowego z nasyceniem

•2.31• Jak rozpoznaje się znak sprzężenia zwrotnego?

Znak sprzężenia zwrotnego można rozpoznać badając obieg sygnałów w zamkniętej pętli sprzężenia. W tym celu należy przerwać (naprawdę lub hipotetycznie) pętlę sprzężenia zwrotnego w dowolnym miejscu i porównać znaki sygnałów na „wejściu” przerwanej pętli i na jej „wyjściu” (tj. w punktach powstałych po przerwaniu pętli). Jeśli znaki te są zgodne — sprzężenie zwrotne jest **dodatnie**, jeśli znaki są przeciwne — sprzężenie zwrotne jest **ujemne**. Najwygodniej jest dokonywać przerywania pętli sprzężenia zwrotnego przy węźle sumującym, zwłaszcza jeśli sprzężenie zwrotne jest bezpośrednie (funkcja f_2 na rys. 2.28A (c) jest tożsamością). Jeśli w takiej sytuacji „wzmocnienie w pętli otwartej” (czyli iloraz $y_2/u_1 = y/u_1$) jest dodatnie, to znakiem sprzężenia zwrotnego jest znak, z jakim gałąź sprzężenia zwrotnego jest dołączona do węzła sumującego. Jeśli jednak wzmocnienie w pętli otwartej jest ujemne (co jest częstym przypadkiem), to znak sprzężenia zwrotnego jest przeciwny do znaku w węźle sumującym. Jeśli znaki odpowiednich sygnałów zmieniają się, np. zależnie od wartości sygnałów lub kształtu przebiegów, podany sposób oceny znaku sprzężenia zwrotnego nie daje jednoznacznej odpowiedzi.

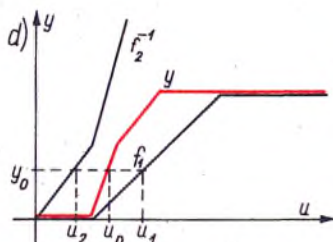
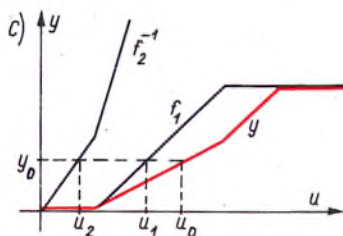
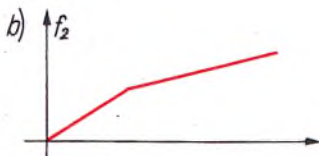
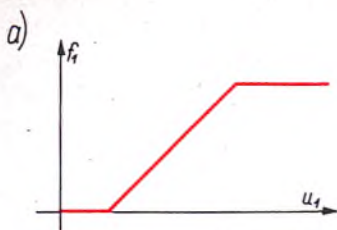
•2.32• Czy można skonstruować graficznie charakterystykę statyczną układu ze sprzężeniem zwrotnym?

Tak, w dość prosty sposób. Przyjmijmy na przykład, że są dane charakterystyki statyczne $f_1(u_1)$ i $f_2(u_2)$ w postaci wykresów — rys. 2.32 (człony f_1 i f_2 mają być połączone zwrotnie, jak na rys. 2.28A (c)). Zbadamy od razu dwa warianty — sprzężenia dodatniego i ujemnego. Konstruujemy charakterystykę wypadkową na podstawie wzoru z pyt. 2.28, mianowicie

$$u = f_1^{-1}(y) \pm f_2(y)$$

przy czym znak „-” odnosi się do sprzężenia dodatniego a „+” — do ujemnego. Zgodnie z tym wzorem można wykreślić charakterystykę $f_1(u_1)$ w zwykłym układzie współrzędnych, natomiast $f_2(y)$ oznacza, że dla tej charakterystyki oś y jest osią zmiennej niezależnej — tzn. charakterystykę $f_2(u_2)$ trzeba wykreślić jako charakterystykę odwrotną $f_2^{-1}(u_2)$ — z odwróceniem osi — patrz rys. 2.18. Następnie należy **odjąć** (dla sprzężenia dodatniego) lub **dodać** (dla sprzężenia ujemnego) **odcięte** obu charakterystyk $f_1(u_1)$ i $f_2^{-1}(u_2)$ przy ustalonej rzędnej y . Punkt o odciętej $u_0 = u_1 \pm u_2$ i rzędnej y_0 jest punktem charakterystyki wypadkowej $y(u)$.

Zauważmy, że dla sprzężenia zwrotnego liniowego ($f_2(u_2)$ — liniowa) mechanizm działania tego sprzężenia będzie polegał na „rozciąganiu” wykresu charakterystyki członu f_1 — w prawo przy sprzężeniu ujemnym, w lewo — przy dodatnim. Kłopotliwą sytuację, omówioną w pyt. 2.28 można



•2.32• Konstrukcja charakterystyk statycznych układu ze sprzężeniem zwrotnym: a) człon $f_1(u_1)$; b) człon w pętli sprzężenia zwrotnego $f_2(u_2)$; c) charakterystyka wypadkowa przy sprzężeniu ujemnym; d) charakterystyka wypadkowa przy sprzężeniu dodatnim

interpretować jako taką, w której wykres charakterystyki wypadkowej staje się „pionowy”.

•2.33• Jakże można podać przykłady połączenia członów na zasadzie sprzężenia zwrotnego?

Rozważmy kilka połączeń członów elementarnych na zasadzie sprzężenia zwrotnego, wskazujących na pewne charakterystyczne cechy wprowadzone do układu przez sprzężenie zwrotne. Przyjmijmy najpierw układ „prawie” liniowy — człon f_1 (rys. 2.28A (c)) wykazujący ograniczenie na poziomie $\pm M$, a w zakresie liniowym charakteryzujący się wzmocnieniem $k_1 > 0$, człon f_2 ściśle liniowy o wzmocnieniu k_2 . Zbadamy wynik takiego połączenia przy współczynniku k_2 zmieniającym się od wartości ujemnych — na dodatnie. Przy $k_2 < 0$ sytuacja jest prosta: w zakresie liniowości członu f_1 otrzymuje się

$$u = \frac{1}{k_1}y + k_2y = \left(\frac{1}{k_1} + k_2\right)y$$

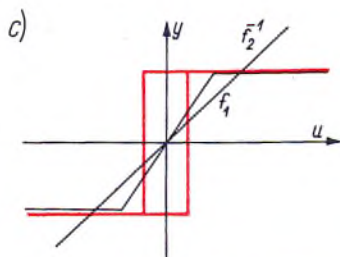
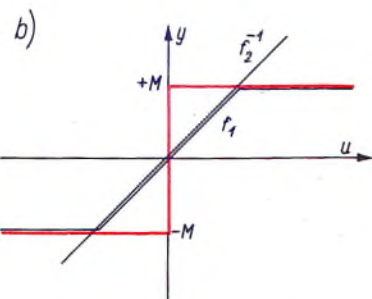
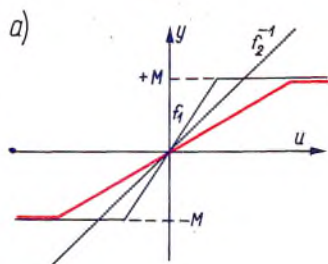
wtedy

$$y = u \frac{k_1}{1 + k_1 k_2}$$

czyli wypadkowy współczynnik wzmocnienia od u do y wynosi

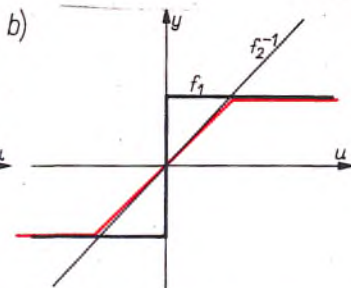
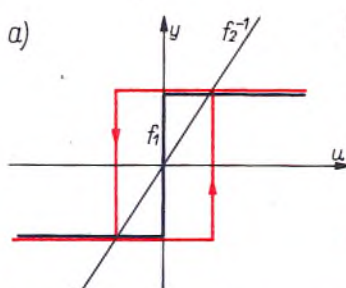
$$k = \frac{k_1}{1 + k_1 k_2}$$

Wypadkowa charakterystyka statyczna jest pokazana na rys. 2.33A (a). Jej kształt jest taki sam jak charakterystyki pierwotnej f_1 , z tym że poziom ograniczenia $\pm M$ pozostaje taki sam, a zakres pracy liniowej, mierzony względem zmiennej wejściowej u , powiększa się $(1 + k_1 k_2)$ -krotnie. Przy $k_2 = 0$ charakterystyka wypadkowa pokrywa się z charakterystyką f_1 . Przy $k_2 > 0$, lecz $k_1 k_2 < 1$, otrzymuje się tylko ilościową różnicę w stosunku do poprzedniej sytuacji — zakres pracy liniowej zmniejsza się, ponieważ czynnik $(1 - k_1 k_2)$ jest mniejszy od jedności. Przy $k_1 k_2 \rightarrow 1$ otrzymujemy sytuację krytyczną omówioną w pyt. 2.28. Konstrukcja graficzna (rys. 2.33A (b)) charakterystyki wypadkowej pokazuje jednak, że dzięki temu, że charakterystyka członu f_1 ma ograniczenie, również charakterystyka wypadkowa jest ograniczona, z tym że krytyczność warunku $k_1 k_2 = 1$ przejawia się w pionowym nachyleniu charakterystyki, a więc nadaniu jej kształtu przekątnikowego. Przy $k_1 k_2 > 1$ najlepiej odwołać się do konstrukcji graficznej (rys. 2.35A (c)), która pokazuje, że charakterystyka wypadkowa jest niejednoznaczna — występuje zjawisko nazwane poprzednio histerezą. Powyższa konstrukcja umożliwia interpretację sensu fizycznego zjawisk prowadzących do histerezy.



•2.23A• Charakterystyki wypadkowe członu z nasyceniem objętego sprzężeniem zwrotnym liniowym: a) sprzężenie ujemne; b) sprzężenie dodatnie „krytyczne”; c) sprzężenie dodatnie większe od krytycznego

•2.23B• Charakterystyki idealnego elementu przekaźnikowego objętego sprzężeniem zwrotnym liniowym: a) dodatnim; b) ujemnym

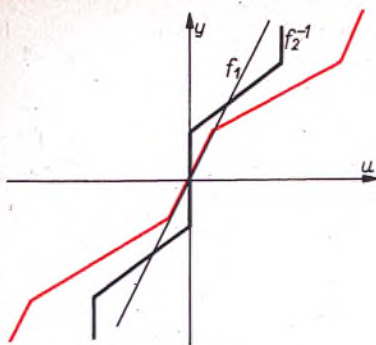


Odwrócimy obecnie sytuację i przyjmijmy, że jako człon f_1 mamy dany idealny element przekaźnikowy, a człon f_2 jest liniowy — jak poprzednio. Przy sprzężeniu zwrotnym dodatnim powinno to wprowadzić efekt histerezy, zaś przy ujemnym prowadzić do „linearyzacji” charakterystyki przekaźnikowej (rys. 2.23B). Zauważmy jednak, że interpretacja drugiego wariantu nie jest łatwa: skoro charakterystyka $f_1(u_1)$ dopuszcza tylko dwie możliwe wartości $y = +M$ i $-M$, to w jaki sposób można otrzymać pośrednie wartości y na charakterystyce wypadkowej? Wyjaśnienie tego na gruncie czysto statycznego nie jest możliwe. Zauważmy, że w „zlinearyzowanym” zakresie charakterystyki wypadkowej obowiązuje warunek $u_1 \approx 0$. Jednak przy $u_1 \approx 0$ wartość y nie jest jednoznaczna. W rzeczywistości występuje wtedy bardzo szybkie przełączanie wartości y pomiędzy $+M$ i $-M$ w taki sposób, że „średnio” otrzymuje się wartość pośrednią $-M < y_{sr} < +M$, przy czym $k_2 y_{sr} = u$. Oporowanie uśrednianiem należy rozumieć w sensie czasowym, uśrednianie w czasie zaś implikuje, że rozważany układ nie jest statyczny, lecz dynamiczny. Jest to niewątpliwie słuszne spostrzeżenie, ponieważ przy dostatecznie szybkim przełączaniu wartości y nie można traktować członu f_2 znajdującego się w pętli sprzężenia zwrotnego jako idealnie statycznego.

Na rysunku 2.23C pokazano jeszcze jeden prosty przykład. Tym razem przyjęto człon f_1 jako liniowy, a człon w pętli f_2 sprzężenia zwrotnego — nieliniowy. Konstrukcja charakterystyki wypadkowej nie budzi wątpliwości. Warto jednak zauważyć, że sprzężenie zwrotne wprowadza do układu tendencję do nadawania charakterystyce wypadkowej kształtu zbliżonego do charakterystyki odwrotnej członu w pętli sprzężenia zwrotnego. Tendencja ta jest tym silniejsza, im większe jest „wzmocnienie” połączenia szeregowego członów f_1 i f_2 .

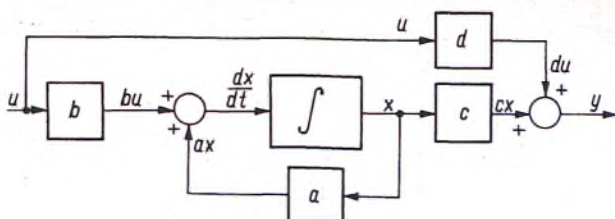
•2.24• Jaką postać ma opis najprostszego układu dynamicznego?

Najprostszym układem dynamicznym jest układ liniowy pierwszego rzędu o jednej zmiennej wejściowej u i jednej zmiennej stanu x . Podstawowym równaniem układu jest **równanie stanu**. Równanie to zazwyczaj przedstawia się jako równanie różniczkowe, wynikające z bilansu zmiennej stanu



•2.33C• Charakterystyka wypadkowa członu liniowego f_1 objętego sprzężeniem nieliniowym f_2

•2.34• Schemat struktury równań układu dynamicznego liniowego pierwszego rzędu



(zmiennej podlegającej akumulacji), a więc uzależniającę prędkość zmian stanu $\frac{dx}{dt}$ od aktualnej wartości zmiennej stanu i od zmiennej wejściowej u . Równanie stanu można przedstawić w postaci

$$\frac{dx(t)}{dt} = ax(t) + bu(t)$$

przy czym współczynniki a i b są stałe (jeśli układ jest stacjonarny). Ponieważ zmienną dostępną dla obserwatora jest wyjście y , więc należy jeszcze podać **równanie wyjścia** uzależniające aktualne wartości y od x i ewentualnie u . Równanie wyjścia jest równaniem algebraicznym, w danym przypadku liniowym

$$y(t) = cx(t) + du(t)$$

Współczynniki a i d mogą być równe zero jako szczególne przypadki opisu ogólnego, natomiast b i c nie. Gdyby było $b = 0$, oznaczałoby to, że rozpatrujemy układ **swobodny** (bez wymuszeń, działający w oderwaniu od otoczenia). Zero-wość współczynnika c oznaczałaby, że nie interesujemy się stanem, a więc dynamiką układu. Najprostszy wariant opisu ma miejsce przy $c = 1$ i $d = 0$ i oznacza, że istotne jest tylko równanie stanu, natomiast wyjście y jest identyczne ze stanem x . Na rysunku 2.34 pokazano schemat struktury powyższych równań.

•2.35• Jak można uogólnić podany sposób opisu na układy wyższych rzędów?

W przypadku układu liniowego rzędu n , lecz o jednym wejściu i wyjściu, równania stanu układu mają postać

$$\frac{dx_1}{dt} = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + b_1u$$

$$\frac{dx_2}{dt} = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + b_2u$$

$$\vdots$$

$$\frac{dx_n}{dt} = a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n + b_nu$$

Równania te można zapisać jako jedno równanie wektorowo-macierzowe

$$\frac{dx}{dt} = Ax + bu$$

przy czym $x = [x_1 \ x_2 \ \dots \ x_n]^T$ jest wektorem stanu (transpozycja podkreśla, że traktujemy ten wektor jako kolumnowy), A jest macierzą kwadratową $n \times n$, której elementami są współczynniki równań stanu, b jest wektorem (kolumnowym) współczynników $b_1, \dots, b_n$ przy funkcji wymuszającej u (skalarnej) w poszczególnych równaniach.

Przy n zmiennych stanu i w przypadku układu liniowego wyjście y jest kombinacją liniową tych zmiennych oraz ewentualnie zmiennej u :

$$y = c_1 x_1 + c_2 x_2 + \dots + c_n x_n + du$$

co można zapisać krócej

$$y = cx + du$$

przy czym $c = [c_1 \ c_2 \ \dots \ c_n]$ jest wektorem wierszowym, zaś d jest liczbą.

Podany sposób opisu nazywa się **opisem w przestrzeni stanów**.

·2.36· Czy opis w przestrzeni stanów jest jednoznaczny dla danego układu?

Nie, ponieważ wybór zmiennych stanu dla danego układu nie jest jednoznaczny. Wprowadzenie przekształcenia **niesobliwego** zmiennych stanu (tj. takiej zamiany współrzędnych w przestrzeni stanów, że przejście od jednych współrzędnych do drugich jest wzajemnie jednoznaczne) spowoduje, że ten sam układ będzie mógł być opisany innymi zmiennymi stanu i równaniami o innej postaci — przy zachowaniu tych samych właściwości. Przekształcenie współrzędnych w przypadku liniowym oznacza, że współrzędne stanu x są kombinacjami liniowymi nowych współrzędnych z (i na odwrót), np.

$$x_1 = p_{11}z_1 + p_{12}z_2 + \dots + p_{1n}z_n$$

$$x_2 = p_{21}z_1 + p_{22}z_2 + \dots + p_{2n}z_n$$

$$\vdots$$

$$x_n = p_{n1}z_1 + p_{n2}z_2 + \dots + p_{nn}z_n$$

Niesobliwość przekształcenia oznacza, że wyznacznik macierzy przekształcenia $P = [p_{ij}]$ jest różny od zera, co zapewni jednoznaczność zamiany zmiennych x na z i na odwrót. Równania można zapisać w skrócie jako

$$x = Pz \quad \text{lub} \quad z = P^{-1}x$$

Zastępując w poprzednich równaniach wektor stanu x przez Pz , otrzymuje się nowe równania

$$\frac{dz}{dt} = P^{-1}APz + P^{-1}bu$$

$$y = cPz + du$$

Tak więc, przy użyciu nowych zmiennych stanu z otrzymuje się równania podobnej postaci, przy czym w miejscu macierzy A pojawia się macierz $P^{-1}AP$, zamiast wektora b — wektor $P^{-1}b$, zamiast wektora c — wektor cP . Najważniejsze jest spostrzeżenie, że macierz $P^{-1}AP$ charakteryzuje cecha podobieństwa do macierzy A (sens tego określenia wyjaśni się dalej). Przesądza to o identyczności właściwości dynamicznych układu niezależnie od wyboru współrzędnych stanu. Pojęciowo jest zresztą oczywiste, że właściwości układu nie mogą zależeć od wyboru zmiennych stanu.

Istnieje nieskończenie wiele sposobów wyboru zmiennych stanu. Ponieważ opisy układu są równoważne bez względu na taki wybór, więc będziemy dążyć do wyboru opisu możliwie najprostszego pod względem rachunkowym.

·2.37· Czy istnieje metoda opisu, która jest jednoznaczna dla danego układu?

Tak, jest to tzw. opis typu **wejście-wyjście**. W opisie tym nie wprowadza się w sposób jawny zmiennych stanu, lecz formułuje się równania uzależniające od razu wejście i wyjście układu. W niektórych przypadkach opis taki wynika w sposób naturalny z przesłanek fizycznych. Punktem wyjścia dla opisu typu wejście-wyjście jest zwykle równanie różniczkowe o postaci

$$a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \dots + a_1 \frac{dy}{dt} + a_0 y = \\ = b_m \frac{d^m u}{dt^m} + \dots + b_1 \frac{du}{dt} + b_0 u$$

Jest to równanie rzędu n względem wyjściowej y , zaś u oznacza wymuszenie (wejście). W przypadkach realizowalnych fizycznie obowiązuje warunek $m \leq n$. Bez utraty ogólności można przyjąć $a_n = 1$, tzn. zredukować wszystkie współczynniki a i b względem a_n .

Należy przypomnieć, że równanie różniczkowe rzędu n można zawsze sprowadzić do układu n równań rzędu pierwszego, wobec czego równanie powyższe po odpowiednim zdefiniowaniu nowych zmiennych (stanu) stałoby się równoważne n równaniom stanu, zaś równanie wyjścia obrazowałoby relację pomiędzy obydwoma rodzajami zmiennych.

Opis typu wejście-wyjście w przedstawionej, pierwotnej postaci nie jest najczęściej zbyt wygodny. Bardzo duże uproszczenie tego opisu uzyskuje się wprowadzając pojęcie **transmitancji operatorowej** oraz pojęcie **całki spłotowej**.

·2.38· Co to jest transmitancja operatorowa?

Załóżmy, że $U(s)$ jest transformatą Laplace'a wymuszenia $u(t)$ pojawiającego się dla $t > 0$, a $Y(s)$ — transformatą szukanego sygnału wyjściowego $y(t)$. Wówczas równanie wejście-wyjście z pyt. 2.37, w dziedzinie transformat, przy zerowych warunkach początkowych, można zapisać jako

$$(s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0)Y(s) = (b_ms^m + \dots + b_1s + b_0)U(s)$$

Transmitancją operatorową (lub krótko transmitancją) nazywa się iloraz transformat wyjścia i wejścia, przy zerowych warunkach początkowych

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_ms^m + \dots + b_1s + b_0}{s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0}$$

Transmitancja $G(s)$ jest funkcją zmiennej zespolonej s i ma tę właściwość, że w wyniku pomnożenia transformaty wejścia $U(s)$ przez transmitancję $G(s)$ otrzymuje się transformatę wyjścia $Y(s)$

$$U(s)G(s) = Y(s)$$

Przebieg czasowy sygnału $y(t)$ w odpowiedzi na doprowadzone wymuszenie $u(t)$ znajdujemy jako transformatę odwrotną

$$y(t) = \mathcal{L}^{-1}\{Y(s)\} = \mathcal{L}^{-1}\{U(s)G(s)\}$$

•2.39• Co to jest transmitancja widmowa?

Transmitancja widmowa powstaje z transmitancji operatorowej $G(s)$ przez podstawienie zamiast zmiennej s — zmiennej $j\omega$, przy czym j jest jednostką urojoną, zaś ω jest liczbą rzeczywistą mającą sens częstotliwości kątowej (pulsacji), mierzonej w radianach na sekundę ($\omega = 2\pi f$, przy czym f oznacza zwykłą częstotliwość przebiegu harmonicznego mierzoną w hercach). A więc

$$G(j\omega) = \frac{b_m(j\omega)^m + \dots + b_1j\omega + b_0}{(j\omega)^n + \dots + a_1j\omega + a_0} = G(s) \Big|_{s=j\omega}$$

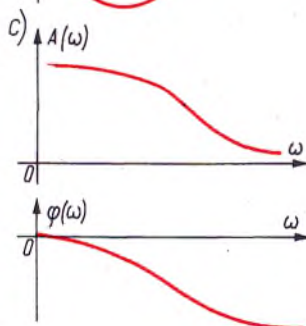
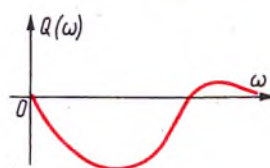
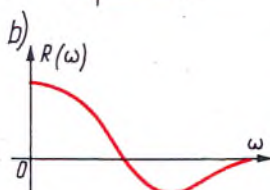
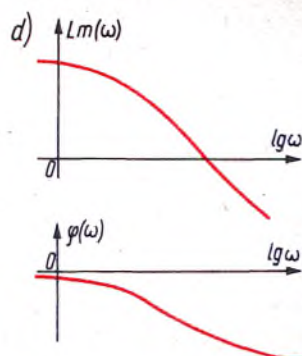
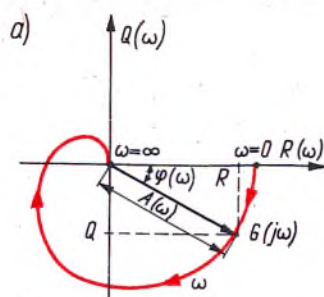
Transmitancja widmowa ma sens „wzmocnienia zespolonego” przebiegu harmonicznego o pulsacji ω (przy czym ω należy traktować jako parametr zmienny). Transmitancję widmową, będącą wielkością zespoloną, można przedstawić w postaci

$$G(j\omega) = A(\omega)e^{j\varphi(\omega)}$$

Odpowiada to sytuacji, w której jeśli przebieg wymuszenia miał postać $u(t) = \sin\omega t$, to przebieg odpowiedzi ma postać $y(t) = A(\omega)\sin[\omega t + \varphi(\omega)]$. Zatem moduł transmitancji widmowej określa wzmocnienie (stosunek amplitud) sygnałów harmonicznym $y(t)$ i $u(t)$, a argument (kąt fazowy) transmitancji widmowej — przesunięcie fazy sygnału $y(t)$ względem $u(t)$.

•2.40• Co to są charakterystyki częstotliwościowe?

Charakterystyki częstotliwościowe otrzymuje się na podstawie transmitancji widmowej, jeśli ω potraktuje się jako zmienną niezależną, przebiegającą wartości od 0 do $+\infty$. Charakterystyki częstotliwościowe mają nie tylko sens pojęcio-



• 2.40 • Charakterystyki częstotliwościowe: a) charakterystyka amplitudowo-fazowa; b) charakterystyki rzeczywista i urojona; c) charakterystyki amplitudowa i fazowa; d) charakterystyki logarytmiczne — amplitudowa i fazowa; e) charakterystyka moduł logarytmiczny-argument

wy, lecz i pomiarowy, ponieważ wprowadzając eksperymentalnie do układu wymuszenie harmoniczne o nastawianej pulsacji $\omega = 2\pi f$ można mierzyć charakterystyki częstotliwościowe na wyjściu układu.

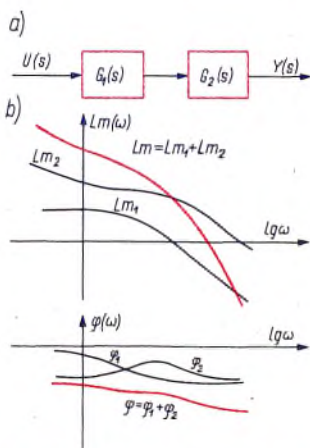
Rozróżnia się następujące charakterystyki częstotliwościowe:

- charakterystykę amplitudową $A(\omega) = |G(j\omega)|$
- charakterystykę fazową $\varphi(\omega) = \arg G(j\omega)$
- charakterystykę rzeczywistą $R(\omega) = \operatorname{Re}[G(j\omega)]$
- charakterystykę urojoną $Q(\omega) = \operatorname{Im}[G(j\omega)]$

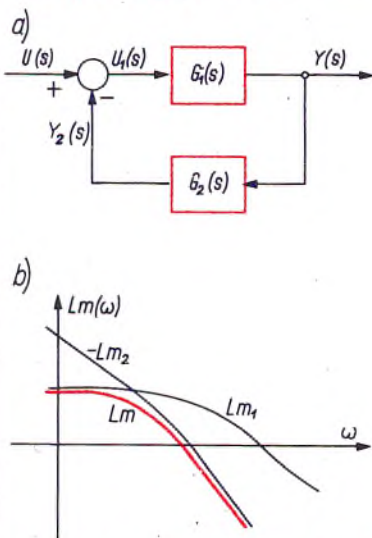
Ponadto określa się charakterystyki częstotliwościowe **logarytmiczne**, jeśli operuje się wykresami. Oś ω i $A(\omega)$ skaluje się logarytmicznie, wprowadzając tzw. moduł logarytmiczny $Lm(\omega) = 20 \lg A(\omega)$, którego jednostką jest decybel (dB); wzmocnieniu 10-krotnemu odpowiada 20 dB, 1-krotnemu — 0 dB. Dla charakterystyki fazowej oś ω skaluje się logarytmicznie, zaś $\varphi(\omega)$ pozostawia się w mierze liniowej.

Jak widać, charakterystyki częstotliwościowe są zestawione parami. Prowadzi to do kilku typowych sposobów przedstawiania graficznego w postaci wykresów. Różne rodzaje tych wykresów pokazano na rys. 2.40. Rysunek 2.40a stanowi elementarną interpretację $G(j\omega)$ na płaszczyźnie zmiennej zespolonej o osiach $R(\omega)$ i $Q(\omega)$. Uzmieniając ω otrzymuje się krzywą $G(j\omega)$ jako tzw. **wykres Nyquista lub charakterystykę amplitudowo-fazową**, ponieważ współrzędne biegunowe każdego punktu na wykresie wyrażają $A(\omega)$ i $\varphi(\omega)$. Charakterystyki $R(\omega)$ i $Q(\omega)$, takie jak na rys. 2.40b, są rzadziej stosowane, jednakże wiele specjalizowanych urządzeń pomiarowych podaje wyniki właśnie w tej postaci. Również charakterystyki $A(\omega)$ i $\varphi(\omega)$ w skali liniowej (rys. 2.40c) są rzadko stosowane. Bardzo powszechnie natomiast stosuje się sposób przedstawienia z rys. 2.40d — są to tzw. **wykresy Bodego**, czyli po prostu charakterystyki logarytmiczne. Wreszcie na rys. 2.40e pokazano tzw. **wykres Blacka** stanowiący połączenie pary charakterystyk moduł logarytmiczny-argument przy pulsacji ω traktowanej jako parametr wykresu.

•2.41• Jakie korzyści daje stosowanie transmitancji i charakterystyk częstotliwościowych?



•2.41A• Sposób tworzenia charakterystyk wypadkowych szeregowego połączenia członów: a) schemat; b) charakterystyki logarytmiczne



•2.41B• Przybliżona ocena kształtu charakterystyk układu ze sprzężeniem zwrotnym: a) schemat układu; b) charakterystyki amplitudowe logarytmiczne

Najważniejsze korzyści wynikają z możliwości znacznego uproszczenia przekształceń przy rozpatrywaniu bardziej złożonych struktur układów. Przekształcenia takie są podobne do przekształceń układów statycznych liniowych.

Największe uproszczenie uzyskuje się przy rozpatrywaniu szeregowego połączenia członów (rys. 2.41A), np. o transmitancjach $G_1(s)$ i $G_2(s)$, bowiem wtedy transmitancja wypadkowa takiego połączenia ma postać $G(s) = G_1(s)G_2(s)$. Nadzwyczaj istotne jest jednak to, że charakterystyki częstotliwościowe logarytmiczne podlegają w takim przypadku po prostu dodawaniu: $Lm(\omega) = Lm_1(\omega) + Lm_2(\omega)$ — ponieważ logarytm iloczynu jest sumą logarytmów; $\varphi(\omega) = \varphi_1(\omega) + \varphi_2(\omega)$ — ponieważ argument iloczynu liczb zespolonych jest sumą argumentów poszczególnych czynników. Stąd właśnie wynika powszechność stosowania wykresów Bodego i również charakterystyk moduł-argument.

Przy równoległym połączeniu członów zależność analityczna jest prosta: $G(s) = G_1(s) + G_2(s)$, jednak nie daje to prawie żadnych korzyści przy określaniu charakterystyk częstotliwościowych. Jedynie w przypadku, gdy w pewnych przedziałach ω jeden ze składników wyraźnie dominuje nad drugimi, charakterystyki częstotliwościowe pokrywają się w przybliżeniu z charakterystyką aktualnie dominującą. Najwięcej trudności przedstawia połączenie ze sprzężeniem zwrotnym (rys. 2.41B). Ponieważ $U_1(s) = U(s) \pm Y_2(s)$, otrzymujemy $Y(s) = G_1(s)U(s) \pm G_1(s)G_2(s)Y(s)$, skąd wynika wzór

$$G(s) = \frac{Y(s)}{U(s)} = \frac{G_1(s)}{1 \pm G_1(s)G_2(s)}$$

przy czym znak „-” obowiązuje dla dodatniego sprzężenia zwrotnego, a „+” — dla ujemnego. Przybliżone oszacowanie przebiegu charakterystyki amplitudowej można uzyskać na podstawie zależności

$$\frac{1}{G(s)} = \frac{1}{G_1(s)} + G_2(s)$$

z której wynika, że jeśli charakterystyki amplitudowe $A_1(\omega)$ i $\frac{1}{A_2(\omega)}$ różnią się znacznie w pewnych zakresach ω , to charakterystyka wypadkowa $A(\omega)$ pokrywa się w przybliżeniu z przebiegiem mniejszego ze składników $A_1(\omega)$ lub $\frac{1}{A_2(\omega)}$. Na rysunku 2.41B pokazano przykład zastosowania tej zasady, przy czym charakterystyce $A_1(\omega)$ odpowiada $Lm_1(\omega)$, a charakterystyce $1/A_2(\omega)$ odpowiada $-Lm_2(\omega)$.

·2.42· Co to są charakterystyki czasowe układu?

Jeśli wymuszenie $u(t)$ ma postać bardzo wąskiego i bardzo wysokiego impulsu o powierzchni jednostkowej, to można taki impuls uznać za przybliżenie tzw. funkcji delta Diraca $\delta(t)$, której transformata Laplace'a jest równa jedności. Wówczas

$$Y(s) \approx G(s) \cdot 1$$

oraz

$$y(t) \approx \mathcal{L}^{-1}\{G(s)\} = g(t)$$

Transformata odwrotna transmitancji $G(s)$, oznaczana zwykle przez $g(t)$, jest to **charakterystyka impulsowa** (lub odpowiedź impulsowa), czyli przebieg uzyskany na wyjściu układu po doprowadzeniu jako wymuszenie impulsu Diraca. Dla bardziej rzeczywistego wymuszenia w postaci funkcji **skoku jednostkowego**

$$1(t) = \begin{cases} 0 & \text{dla } t < 0 \\ 1 & \text{dla } t > 0 \end{cases}$$

$$\text{o transformacie } \mathcal{L}\{1(t)\} = \frac{1}{s}$$

otrzymujemy $Y(s) = \frac{G(s)}{s}$, a więc

$$y(t) = \mathcal{L}^{-1}\left\{\frac{G(s)}{s}\right\} = \int_0^t g(\tau) d\tau = h(t)$$

Jest to **charakterystyka skokowa** układu (lub odpowiedź skokowa, jeśli przyjąć ogólnie $u(t) = A \cdot 1(t)$). Obydwie funkcje $g(t)$ i $h(t)$ noszą nazwę charakterystyk czasowych, przy czym istotna jest ich zależność

$$h(t) = \int_0^t g(\tau) d\tau; \quad g(t) = \frac{d}{dt} h(t)$$

a także ich związek z transmitancją.

Charakterystyki czasowe (podobnie jak częstotliwościowe) można określać doświadczalnie (w przybliżeniu), a zatem mogą stanowić podstawę do identyfikacji układu. Wyznaczają one jednoznaczny opis typu wejście-wyjście.

2.43. Co to jest opis za pomocą całki spłotowej?

Jest to jeszcze jedna postać opisu typu wejście-wyjście, mająca cechę jednoznaczności dla danego układu. Opis ten wynika wprost z właściwości przekształcenia Laplace'a. Odpowiednikiem iloczynu transformat $U(s)G(s) = Y(s)$ w dziedzinie czasu jest bowiem tzw. iloczyn spłotowy lub całka spłotowa

$$y(t) = \int_{-\infty}^{\infty} u(\tau)g(t-\tau)d\tau = \int_{-\infty}^{\infty} g(\tau)u(t-\tau)d\tau$$

Zmienna τ jest zmienną całkowania. Opis ten jest podobny do opisu za pomocą transmitancji pod tym względem, że wymaga przeprowadzenia pewnej operacji nad funkcją $g(t)$ charakteryzującą układ i funkcją $u(t)$ reprezentującą wymuszenie; operacja ta odbywa się całkowicie w dziedzinie czasu. W realnej sytuacji całkę spłotową liczy się w granicach skończonych, ponieważ wymuszenie $u(t)$ ma sens tylko dla $t \geq 0$, zaś przy $u(t) = 0$ dla $t < 0$ również charakterystyka impulsowa $g(t)$ jest równa 0 dla $t < 0$, jeśli układ jest przyczynowy (tzn. nie wykazuje reakcji, zanim nie zostanie pobudzony). Wobec tego $u(\tau) = 0$ dla $\tau < 0$ oraz $g(t-\tau) = 0$ dla $\tau > t$, a więc „realne” granice całkowania są następujące:

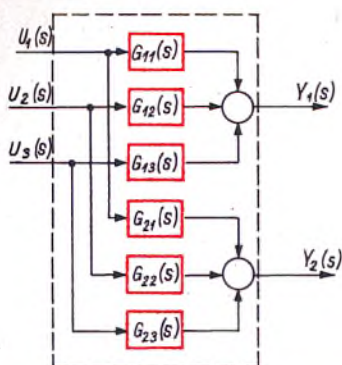
$$y(t) = \int_0^t u(\tau)g(t-\tau)d\tau = \int_0^t g(\tau)u(t-\tau)d\tau$$

2.44. Jaką postać ma zapis wektorowo-macierzowy w przypadku układu o wielu wejściach i wyjściach?

W ogólnym przypadku układu liniowego stacjonarnego n -tego rzędu, o r wejściach i m wyjściach, równania układu można zapisać następująco:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{1r} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{nr} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_r \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} d_{11} & \dots & d_{1r} \\ \vdots & & \vdots \\ d_{m1} & \dots & d_{mr} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_r \end{bmatrix}$$



·2.44· Struktura układu dynamicznego o 3 wejściach i 2 wyjściach, opisanego transmitancją macierzową

$$G(s) = \begin{bmatrix} G_{11}(s) & G_{12}(s) & G_{13}(s) \\ G_{21}(s) & G_{22}(s) & G_{23}(s) \end{bmatrix}$$

gdzie $\dot{x}$ oznacza pochodną x względem czasu, lub w zwartym zapisie wektorowo-macierzowym

$$\dot{x} = Ax + Bu$$

$$y = Cx + Du$$

przy czym x jest wektorem n -wymiarowym, u — wektorem r -wymiarowym, y — wektorem m -wymiarowym, macierze współczynników mają wymiary odpowiednio: $A - n \times n$ (macierz kwadratowa), $B - n \times r$, $C - m \times n$, $D - m \times r$. W przedstawionym zapisie mieszczą się wszystkie przypadki, włącznie z najprostszym z pyt. 2.34, gdzie wszystkie wielkości były skalarami (x, a, b, u, y), w miejscu macierzy C znajdowała się 1, zaś macierz D była zerowa. Zapis wektorowo-macierzowy można także rozszerzyć na opis typu wejście-wyjście. Można bowiem określić **transmitancję macierzową** jako macierz, której elementami są transmitancje łączące poszczególne wejścia z poszczególnymi wyjściami w następujący sposób:

$$Y(s) = \begin{bmatrix} Y_1(s) \\ Y_2(s) \\ \vdots \\ Y_m(s) \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) & \dots & G_{1r}(s) \\ G_{21}(s) & G_{22}(s) & \dots & G_{2r}(s) \\ \vdots & \vdots & & \vdots \\ G_{m1}(s) & G_{m2}(s) & \dots & G_{mr}(s) \end{bmatrix} \begin{bmatrix} U_1(s) \\ U_2(s) \\ \vdots \\ U_r(s) \end{bmatrix} = G(s)U(s)$$

przy czym $Y(s)$ jest m -wymiarowym wektorem transformat wyjść, $U(s)$ jest r -wymiarowym wektorem transformat wejść, zaś $G(s)$ — transmitancją macierzową o wymiarach $m \times r$. Ogólnie $m \neq r$, więc $G(s)$ jest macierzą prostokątną.

Przez analogię można określić **macierz charakterystyk impulsowych** jako transformatę odwrotną transmitancji macierzowej

$$g(t) = \mathcal{L}^{-1}\{G(s)\} = \begin{bmatrix} g_{11}(t) & g_{12}(t) & \dots & g_{1r}(t) \\ g_{21}(t) & g_{22}(t) & \dots & g_{2r}(t) \\ \vdots & \vdots & & \vdots \\ g_{m1}(t) & g_{m2}(t) & \dots & g_{mr}(t) \end{bmatrix}$$

Na rysunku 2.44 pokazano strukturę układu o 3 wejściach i 2 wyjściach. Już przy tak niewielkich liczbach wejść i wyjść widać, że zapis macierzowy wprowadza ogromną zwięźłość sformułowań przy zachowaniu pełnej systematyczności opisu.

·2.45· Czy istnieje ścisły związek pomiędzy opisem w przestrzeni stanów a transmitancją?

Tak, przy czym należy pamiętać, że tylko opis transmitancyjny jest jednoznaczny, a więc wielu różnym postaciom opisu w przestrzeni stanów może odpowiadać jeden opis transmitancyjny.

W najprostszym przypadku (skalarnym) związek obydwu

metod opisu otrzymuje się bardzo łatwo. Stosując przekształcenie Laplace'a do równania stanu o postaci

$$\dot{x} = ax + bu$$

otrzymujemy (przy zerowych warunkach początkowych)

$$sX(s) = aX(s) + bU(s)$$

stąd

$$X(s) = \frac{b}{s-a} U(s)$$

oraz, na podstawie równania wyjścia $y = cx + du$,

$$Y(s) = \left(\frac{cb}{s-a} + d \right) U(s)$$

A więc transmitancja układu jest następująca:

$$G(s) = \frac{cb}{s-a} + d$$

W przypadku układu rzędu n , o jednym wejściu i jednym wyjściu, przekształcenia wymagają zastosowania algebry macierzy. Równanie stanu

$$\dot{x} = Ax + bu$$

po transformacji przyjmuje postać

$$sX(s) = AX(s) + bU(s)$$

Stąd

$$(sI - A)X(s) = bU(s)$$

(I oznacza macierz jednostkową o wymiarze n).

Transformatę wektora stanu można zapisać jako

$$X(s) = (sI - A)^{-1} b U(s)$$

zatem

$$Y(s) = [c(sI - A)^{-1} b + d] U(s)$$

i

$$G(s) = c(sI - A)^{-1} b + d$$

Jest to wyrażenie skalarne, ponieważ człon $c(sI - A)^{-1} b$ jest iloczynem wektora wierszowego (macierzy $1 \times n$) i macierzy kwadratowej $n \times n$ oraz wektora kolumnowego (macierzy $n \times 1$).

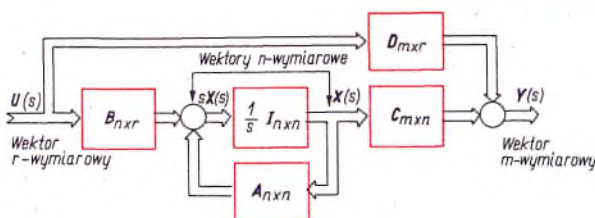
Powyższe zależności można łatwo rozszerzyć na przypadek układu o n zmiennych stanu, r wejściach i m wyjściach. Rozszerzenie polega na zastąpieniu wektora b macierzą B , wektora wierszowego c — macierzą C i skalarą d — macierzą D . W związku z tym transformaty wektora stanu i wektora wyjść mają postacie

$$X(s) = (sI - A)^{-1} B U(s)$$

$$Y(s) = [C(sI - A)^{-1} B + D] U(s)$$

przy czym $U(s)$ jest transformatą wektora wymuszeń.

2.45. Struktura układu dynamicznego n -tego rzędu o r wejściach i m wyjściach



Stąd

$$G(s) = C(sI - A)^{-1}B + D$$

Powyższy model typu wejście-wyjście, sformułowany za pomocą macierzy A , B , C i D , można poglądowo przedstawić w sposób strukturalny. Na rysunku 2.34 pokazano już najprostszy wariant tego modelu przy $u = r = m = 1$ (wszystkie zmienne są skalarami). Rysunek 2.45 przedstawia wariant ogólny, o budowie analogicznej, z tym jednak, że w najważniejszej części układu występuje n integratorów związanych z poszczególnymi zmiennymi stanu (co symbolicznie zaznaczono jako $\frac{1}{s}I$), macierz A obrazuje wszystkie połączenia „skrośne” pomiędzy poszczególnymi zmiennymi stanu i pochodnymi zmiennych stanu na wejściach integratorów, natomiast macierze B , C i D wyrażają połączenia typu „każdy z każdym” odpowiednich zmiennych.

Porównując postacie transmitancji skalarnej $G(s)$ z poprzedniego pytania i z pyt. 2.38, można dokonać kilku spostrzeżeń. Jeśli transmitancja ma mieć postać ilorazu dwóch wielomianów zmiennej s , to, po pierwsze, współczynnik d różny od zera można otrzymać przez podzielenie wielomianów licznika i mianownika, wyłącznie przy równych stopniach tych wielomianów. Po wyłączeniu składnika d (reprezentującego statyczną relację pomiędzy wejściem i wyjściem) pozostaje część dynamiczna, w której podstawową rolę pełni człon $(sI - A)^{-1}$. Jest to macierz o wymiarach $n \times n$, której wszystkie elementy są dzielone przez wyznacznik

2.46. Co to jest wielomian charakterystyczny układu?

Wyznacznik macierzy $(sI - A)$ jest wielomianem stopnia n , który identyfikujemy z wielomianem stopnia n występującym w mianowniku transmitancji

$$\det(sI - A)$$

Jest to właśnie wielomian charakterystyczny układu, a zarazem wielomian charakterystyczny macierzy A . Stopień wielomianu charakterystycznego, równy n , jest oczywiście równy rzędowi układu dynamicznego. Tak więc macierz A i wielomian charakterystyczny pełnią najważniejszą rolę

$$G(s) = \frac{b_n s^m + \dots + b_1 s + b_0}{s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0}$$

Jest to właśnie wielomian charakterystyczny układu, a zarazem wielomian charakterystyczny macierzy A . Stopień wielomianu charakterystycznego, równy n , jest oczywiście równy rzędowi układu dynamicznego. Tak więc macierz A i wielomian charakterystyczny pełnią najważniejszą rolę

przy określeniu właściwości dynamicznych układu. Od współczynników b i c zależy postać transmitancji, zwłaszcza jej licznika, natomiast nie mają one wpływu na wielomian charakterystyczny.

•2.47• Co to są wartości własne układu?

Są to miejsca zerowe wielomianu charakterystycznego układu czyli pierwiastki **równania charakterystycznego** mającego postać

$$\det(sI - A) = s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0 = 0$$

Wartości własne układu są zarazem wartościami własnymi macierzy A , co wynika z powyższego określenia. Tylko wśród wartości własnych mogą się znaleźć bieguny transmitancji operatorowej, tzn. miejsca zerowe mianownika transmitancji, przy których transmitancja przestaje być określona. Ważną cechą wartości własnych macierzy jest to, że są one niezmiennicze dla klasy macierzy podobnych (cecha podobieństwa łączy np. macierz A ze wszystkimi macierzami typu $P^{-1}AP$). Potwierdza to spostrzeżenie z pyt. 2.36, że nieosobliwe przekształcenie współrzędnych stanu nie zmienia zasadniczych właściwości układu. Właściwości te są w istocie określone przez wartości własne, które są **jednoznaczne** w odróżnieniu od współrzędnych stanu i macierzy A układu.

Układ rzędu n ma oczywiście n wartości własnych $s_1, s_2, \dots, s_n$. Mogą występować wartości własne wielokrotnie (w skrajnym przypadku układ może mieć n identycznych wartości własnych). Wartości własne mogą być liczbami rzeczywistymi lub zespolonymi tworzącymi pary sprzężone.

•2.48• Jaki związek mają wartości własne z przebiegami dynamicznymi w układzie?

Związek ten ma zasadnicze znaczenie. Najlepiej można to pokazać rozpatrując układ dynamiczny **swobodny** (bez wymuszeń). Równanie stanu układu swobodnego ma postać (w zapisie wektorowo-macierzowym)

$$\dot{x} = Ax$$

Rozwiązanie tego równania przy zerowych warunkach początkowych jest trywialne ($x = 0$). Natomiast przyjęcie warunków początkowych różnych od zera daje w efekcie tzw. **rozwiązanie swobodne**. W przypadku układu pierwszego rzędu, tzn. przy równaniu

$$\dot{x} = ax; \quad x(0) = x^0$$

rozwiązanie swobodne ma oczywiście postać

$$x(t) = e^{at} x^0$$

(co łatwo sprawdzić i co wynika z metody ogólnej rozwiązywania równań różniczkowych liniowych o stałych współczynnikach). W przypadku układu rzędu n rozwiązanie rów-

niania $\dot{x} = Ax$ przy $x(0) = x^0$ można zapisać w analogicznej postaci

$$x(t) = e^{At}x^0$$

Funkcja wykładnicza macierzy kwadratowej jest również macierzą kwadratową i jest określona jako szereg nieskończony analogiczny do przypadku skalarnego: $e^{At} = I + At + \frac{1}{2}A^2t^2 + \frac{1}{6}A^3t^3 + \dots$. Istotne są związki: $e^{At} = I$ dla $t = 0$

oraz $\frac{d}{dt} e^{At} = Ae^{At}$. Wyrażenie e^{At} jest to tzw. **macierz podstawowa** rozwiązań układu dynamicznego liniowego. Macierz tę można obliczyć wykorzystując przekształcenie Laplace'a. Transformując równanie $\dot{x} = Ax$ otrzymujemy $sX(s) - x^0 = AX(s)$

wobec tego transformatę rozwiązania swobodnego można wyrazić jako

$$X(s) = (sI - A)^{-1}x^0$$

A więc macierz e^{At} jest transformatą odwrotną macierzy $(sI - A)^{-1}$. Wspólnym mianownikiem wszystkich elementów macierzy $(sI - A)^{-1}$ jest oczywiście wyznacznik $\det(sI - A)$ czyli wielomian charakterystyczny, który można zapisać w postaci rozłożonej na czynniki

$$\det(sI - A) = (s - s_1)(s - s_2) \dots (s - s_n)$$

Wobec tego transformata odwrotna macierzy $(sI - A)^{-1}$ powinna zawierać wśród swoich elementów wszystkie wyrażenia o postaci $e^{s_1 t}$, $e^{s_2 t}$, ..., $e^{s_n t}$.

Wynika stąd ogólny wniosek: tak jak dla układu pierwszego rzędu, gdzie w rozwiązaniu swobodnym pojawił się wyraz e^{at} (przy czym a było wartością własną), również w rozwiązaniu swobodnym układu n -tego rzędu pojawiają się wyrażenia typu $e^{s_i t}$ ($i = 1, \dots, n$), przy czym s_i są wartościami własnymi. A więc wartości własne determinują możliwą postać rozwiązań swobodnych układu dynamicznego.

•2.49. Jaki jest związek macierzy podstawowej z rozwiązaniami wymuszonymi?

Rozważmy najpierw najprostszy przypadek układu pierwszego rzędu z wymuszeniem, przy zerowym warunku początkowym. Równanie stanu ma postać

$$\dot{x} = ax + bu; \quad x(0) = 0$$

Wykorzystując przekształcenie Laplace'a można napisać $sX(s) = aX(s) + bU(s)$

$$X(s) = \frac{1}{s-a} b U(s)$$

Ponieważ $\mathcal{L}^{-1} \left\{ \frac{1}{s-a} \right\} = e^{at}$, więc $x(t)$ można wyrazić za pomocą całki spłotowej

$$x(t) = \int_0^t e^{A(t-\tau)} b u(\tau) d\tau$$

Analogicznie, w przypadku układu rzędu n równanie stanu układu z wymuszeniem można zapisać jako

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u; \quad \mathbf{x}(0) = \mathbf{0}$$

lub w dziedzinie transformat

$$s\mathbf{X}(s) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}U(s)$$

stąd

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}U(s)$$

Transformatą odwrotną macierzy $(s\mathbf{I} - \mathbf{A})^{-1}$ jest macierz podstawowa e^{At} , więc również w tym przypadku rozwiązanie $\mathbf{x}(t)$ można przedstawić za pomocą spłotu

$$\mathbf{x}(t) = \int_0^t e^{A(t-\tau)} \mathbf{B}u(\tau) d\tau$$

Otrzymujemy w ten sposób ciekawą interpretację macierzy podstawowej e^{At} lub w dziedzinie transformat macierzy $(s\mathbf{I} - \mathbf{A})^{-1}$. Wyrażenia te (z uwzględnieniem macierzy współczynników $\mathbf{B}$) pełnią rolę „charakterystyki impulsowej” lub „transmitancji” w odniesieniu do relacji wejście-stan. Przejście od relacji wejście-stan do relacji wejście-wyjście jest zabiegiem stosunkowo prostym, wymagającym tylko przekształceń algebraicznych. Ogólne równanie wyjścia zapisane za pomocą spłotu ma postać

$$y(t) = \int_0^t \mathbf{C}e^{A(t-\tau)} \mathbf{B}u(\tau) d\tau + \mathbf{D}u(t)$$

Jest to odpowiednik czasowy zależności transmitacyjnej (pyt. 2.45) pomiędzy wejściem i wyjściem. Wyrażenie to odpowiada dokładnie całce spłotowej, przy czym macierz charakterystyk impulsowych ma postać

$$\mathbf{g}(t) = \mathbf{C}e^{At}\mathbf{B} + \mathbf{D}\delta(t)$$

co jest zgodne z transformatą odwrotną transmitancji macierzowej $\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$.

Obecność macierzy e^{At} w wyrażeniu na rozwiązanie wymuszone powoduje, że w tym rozwiązaniu mogą pojawić się charakterystyczne człony o postaci $e^{s_i t}$. Ponownie więc potwierdza się fakt, że wartości własne s_i ($i = 1, \dots, n$) dostarczają najważniejszej informacji o właściwościach dynamicznych układu. Należy podkreślić, że wartości własne decydują o postaci przebiegów **przejściowych** w układzie, w odróżnieniu od przebiegów **ustalonych**, mających najczęściej postać taką samą jak wymuszenie. W najbardziej ogólnym przypadku przy niezerowych warunkach początkowych i przy obecności wymuszenia, rozwiązanie jest sumą składowej swobodnej i wymuszonej, przy czym przebieg prze-

ściowy obejmuje całą składową swobodną i część składowej wymuszonej, zawierającą składniki $e^{s_i t}$, których pojawienie się jest spowodowane przez wymuszenie. Można to zapisać jako

$$y(t) = C \left\{ e^{A t} x^0 + \int_0^t e^{A(t-\tau)} B u(\tau) d\tau \right\} + D u(t)$$

2.50. Co oznacza pojęcie stabilności układu?

Istnieje wiele interpretacji pojęcia stabilności. W przypadku układów liniowych stacjonarnych o parametrach skupionych intuicyjne rozumienie stabilności jest całkowicie ścisłe. Intuicyjnie za układ stabilny uważa się taki, którego rozwiązanie swobodne (przy niezerowych warunkach początkowych) pozostaje ograniczone w dowolnym czasie. Jest to równoważne żądaniu, aby przy wymuszeniu ograniczonym co do wartości i czasu trwania odpowiedź układu była również ograniczona. Pewne uściślenie wprowadza pojęcie **stabilności asymptotycznej**. Stabilność asymptotyczna oznacza, że układ nie tylko jest stabilny, a więc jego rozwiązania są ograniczone, ale przy czasie dążącym do nieskończoności rozwiązanie swobodne dąży do zera. Odpowiada to żądaniu, aby rozwiązanie wymuszone było ograniczone nawet przy wymuszeniu (ograniczonym) trwającym dowolnie długo.

2.51. Czy stabilność układu liniowego zależy od wartości własnych układu?

Tak, w układzie liniowym stacjonarnym istnieje jednoznaczny związek pomiędzy wartościami własnymi a stabilnością. Ponieważ w rozwiązaniu równania stanu pojawiają się składniki zawierające wyrazy $e^{s_i t}$, więc układ jest stabilny, jeśli wszystkie wartości własne s_i ($i = 1, \dots, n$) mają ujemne części rzeczywiste

$$\operatorname{Re} s_i < 0; \quad i = 1, \dots, n$$

Jeśli choć jedna wartość własna ma dodatnią część rzeczywistą, to układ jest niestabilny. Jeśli występują wartości własne o zerowych częściach rzeczywistych (a więc rzeczywiste zerowe lub czysto urojone), to układ pozostaje stabilny, jeśli te wartości są pojedyncze. Wynika to z ograniczoności wyrażen $e^{0 \cdot t} = 1$ lub $|e^{j\omega t}| = 1$. Nie zachodzi tu jednak warunek stabilności asymptotycznej. Przykładami członów stabilnych nieasymptotycznie są człon całkujący (transmitancja $\frac{1}{s}$, wartość własna $s_1 = 0$, rozwiązanie swobodne jest stałe, $x(t) = x^0$, rozwiązanie wymuszone pozostaje ograniczone i stałe, jeśli wymuszenie $u(t)$ w skończonym czasie zaniknie do zera) lub idealny człon oscylacyjny (transmitancja $\frac{1}{s^2 + \omega^2}$, wartości własne $\pm j\omega$, rozwiązanie swobod-

ne — sinusoidalne). Jeśli wartość własna przy $\text{Re } s_i = 0$ jest wielokrotna, to w rozwiązaniu pojawiają się człony typu $e^{s_i t}$, $t e^{s_i t}$, $t^2 e^{s_i t}$ itd. w zależności od krotności wartości własnej, co oczywiście powoduje nieograniczoność rozwiązań.

2.52. Jakie są podstawowe człony dynamiczne i ich właściwości?

Przykłady podstawowych członów dynamicznych najlepiej przedstawić graficznie, jak to pokazano w tabl. 2.52, w której zestawiono transmitancje, szkice rozmieszczenia wartości własnych na płaszczyźnie zmiennej zespolonej oraz szkice charakterystyk skokowych i charakterystyk częstotliwościowych. Charakterystyki częstotliwościowe przedstawiono w postaci przybliżonej, asymptotycznej.

Idea charakterystyk asymptotycznych dotyczy charakterystyk logarytmicznych amplitudowych, przy czym korzysta się ze spostrzeżenia, że człon całkujący o transmitancji $\frac{1}{Ts}$

ma charakterystykę amplitudową $A(\omega) = \frac{1}{T\omega}$, stąd charakterystykę logarytmiczną można wyrazić jako $L_m(\omega) = -20 \log T - 20 \log \omega$. Pierwszy składnik jest stały, drugi zaś wskazuje, że charakterystyka logarytmiczna członu całkującego jest linią prostą o nachyleniu -20 dB na dekadę (zmianę 10-krotną) ω .

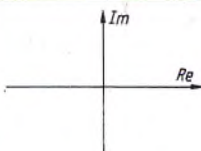
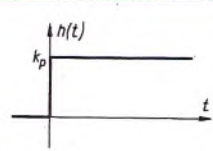
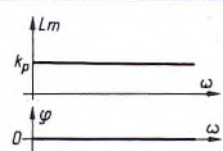
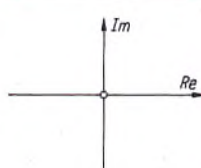
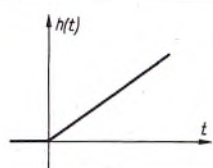
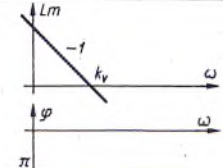
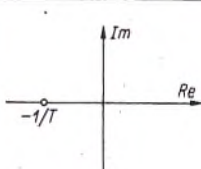
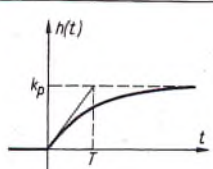
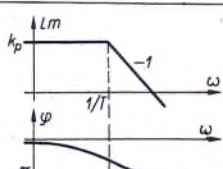
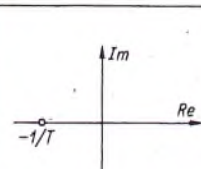
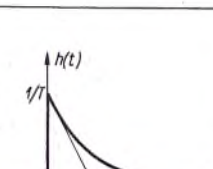
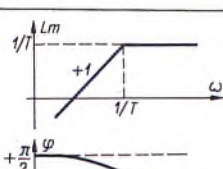
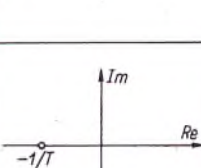
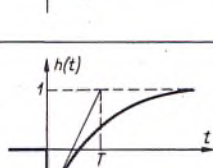
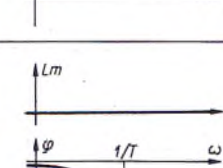
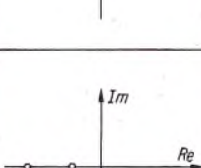
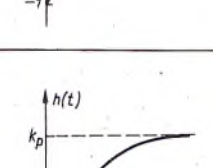
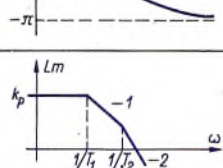
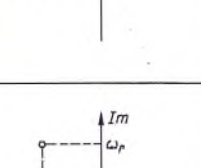
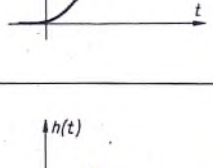
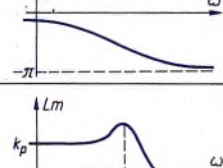
Dla członu proporcjonalnego obowiązuje warunek $L_m(\omega) = \text{const}$, tzn. charakterystyka jest linią prostą poziomą (0 dB/dek). Charakterystykę członu inercyjnego można przybliżyć na podstawie zależności

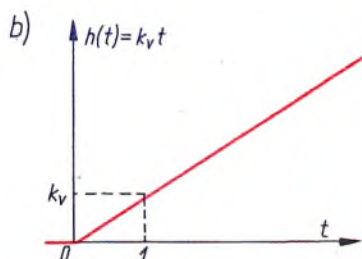
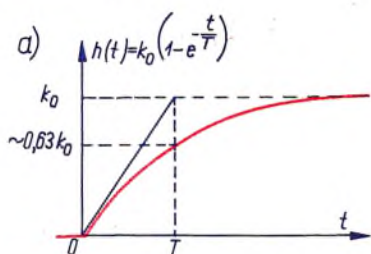
$$A(\omega) = \left| \frac{1}{j\omega T + 1} \right| = \frac{1}{\sqrt{\omega^2 T^2 + 1}} \approx \begin{cases} 1 & \text{dla } \omega \ll \frac{1}{T} \\ \frac{1}{\omega T} & \text{dla } \omega \gg \frac{1}{T} \end{cases}$$

Tak więc otrzymuje się dwie asymptoty: poziomą (0 dB/dek) dla małych ω i asymptotę o nachyleniu -20 dB/dek (jak dla członu całkującego) dla dużych ω . Punkt przecięcia obu asymptot — tzw. **punkt załamania** charakterystyki leży przy $\omega = \frac{1}{T}$; w tym też punkcie występuje największa rozbieżność charakterystyki dokładnej i asymptotycznej, mianowicie $A(\omega) = \frac{1}{\sqrt{2}}$ czyli -3 dB w porównaniu z wartością 1 wynikającą z przybliżeń.

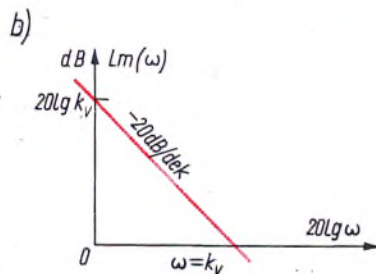
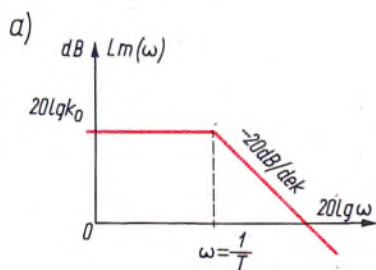
W podobny sposób okazuje się, że dla układu drugiego rzędu charakterystyka logarytmiczna ma nachylenie nie przekraczające -40 dB/dek itd.; jeśli bowiem transmitancja ma postać $\frac{1}{(T_1 s + 1)(T_2 s + 1)}$, to charakterystyka asymptotyczna

2.52. Podstawowe czony dynamiczne

Nazwa czonu	Transmitancja	Wartości własne	Charakterystyka skokowa	Charakterystyka częstotliwościowa
Proporcjonalny	$G(s) = k_p$			
Całkujący	$G(s) = \frac{k_v}{s}$			
Inercyjny	$G(s) = \frac{k_p}{Ts+1}$			
Różniczkujący (rzeczywisty)	$G(s) = \frac{s}{Ts+1}$			
Przesuwnik fazowy	$G(s) = \frac{1-Ts}{1+Ts}$			
Dwuinercyjny	$G(s) = \frac{k_p}{(T_1s+1)(T_2s+1)}$			
Oscylacyjny	$G(s) = \frac{k_p}{T_n^2s^2 + 2\zeta T_ns + 1}$			



• 2.53A • Określanie współczynników wzmocnienia i stałej czasowej: a) charakterystyka skokowa członu inercyjnego pierwszego rzędu; b) członu całkującego



• 2.53B • Określanie parametrów układów z charakterystyk częstotliwościowych asymptotycznych: a) charakterystyka członu inercyjnego $G(s) = \frac{k_0}{Ts+1}$; b) charakterystyka członu całkującego $G(s) = \frac{1}{Ts} = \frac{k_v}{s}$, ($k_v = \frac{1}{T}$)

ma dwa załamania: najpierw dla $\omega = \frac{1}{T_1}$ (jeśli $T_1 > T_2$), po czym obowiązuje nachylenie -20 dB/dek, następnie dla $\omega = \frac{1}{T_2}$, po czym nachylenie zwiększa się o kolejne -20 dB/dek. Ten tok postępowania można dość łatwo uogólnić i wykorzystać przy szkicowaniu charakterystyk logarytmicznych układów dowolnego rzędu. Nachylenie -20 dB/dek określa się często umownie jako „ -1 ”, -40 dB/dek jako „ -2 ” itd. — tak też oznaczono nachylenia charakterystyk w tabl. 2.52.

• 2.53 • Jakie są najważniejsze parametry członów dynamicznych?

Najważniejszymi parametrami członów dynamicznych są współczynniki wzmocnienia, stałe czasowe oraz dla członów oscylacyjnych — częstotliwości własne i współczynniki tłumienia.

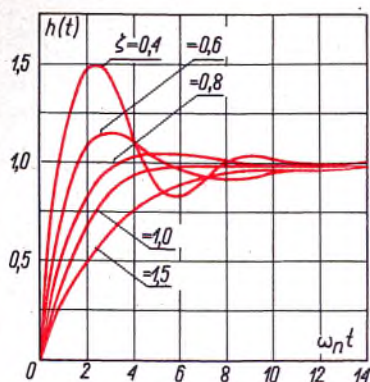
Współczynnikiem wzmocnienia (statycznego) k_0 nazywa się granicę, do której dąży transmitancja przy $s \rightarrow 0$, co odpowiada wzmocnieniu w stanie ustalonym. W przypadku gdy rozpatruje się człon całkujący lub układ zawierający całkowanie, nie można tak określić wzmocnienia, ponieważ wtedy $|G(s)| \rightarrow \infty$ przy $s \rightarrow 0$. W takim przypadku określa się tzw. **współczynnik wzmocnienia prędkościowego** (k_v) jako granicę $sG(s)$ przy $s \rightarrow 0$, co odpowiada stosunkowi prędkości zmian sygnału wyjściowego w stanie ustalonym, przy stałym sygnale wejściowym; k_v ma wymiar odwrotności czasu.

Stalą czasową nazywa się współczynnik T (o wymiarze czasu), pojawiający się w elementarnym członie inercyjnym pierwszego rzędu o transmitancji

$$\frac{k_0}{Ts+1} = \frac{k_0}{T} \frac{1}{s + \frac{1}{T}}$$

Jak widać, równaniem charakterystycznym tego członu jest $s + \frac{1}{T} = 0$, a więc wartość własna $s_0 = -\frac{1}{T}$. Ogólnie, stałe czasowe można utożsamiać z odwrotnościami wartości własnych rzeczywistych. Stałe czasowe określają szybkość możliwych zmian przebiegów rozwiązań, ponieważ określają one składniki typu $e^{s_0 t} = e^{-\frac{t}{T}}$, a więc stała T określa przedział czasu, po którym przebieg zmieni swoją wartość e -krotnie (rys. 2.53A). W przypadku układu wyższego rzędu, mającego kilka stałych czasowych, związek z postacią przebiegu jest bardziej skomplikowany.

Współczynniki wzmocnienia oraz stałe czasowe mają bardzo wyraźną interpretację na charakterystykach amplitudowych (logarytmicznych), co pokazano na rys. 2.53B.



·2.53C· Rodzina charakterystyk skokowych członu oscylacyjnego

Elementarny człon oscylacyjny o transmitancji

$$\frac{1}{T_n^2 s^2 + 2\zeta T_n s + 1} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}; \quad \omega_n = \frac{1}{T_n}$$

ma parę wartości własnych

$$s_{1,2} = -\omega_n(\zeta \pm \sqrt{\zeta^2 - 1})$$

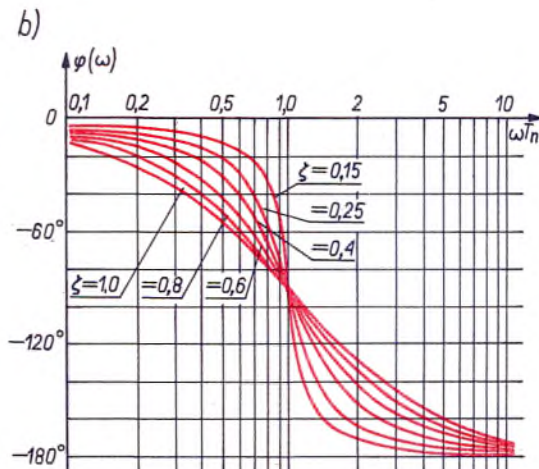
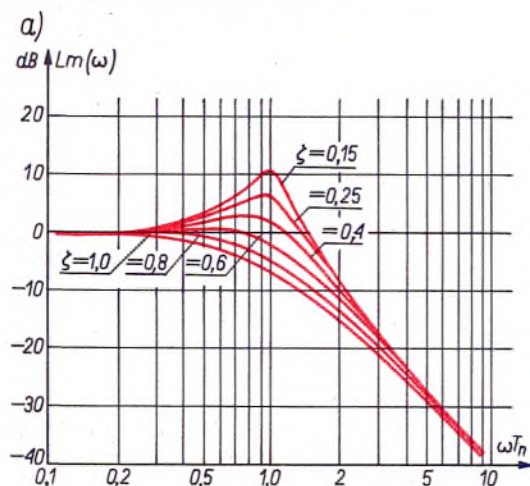
Właściwości istotnie oscylacyjne (a więc wartości własne zespolone) występują przy $\zeta^2 < 1$. Współczynnik ζ jest to **współczynnik tłumienia względnego**, ω_n określa **pulsację drgań własnych** nietłumionych (tzn. przy $\zeta = 0$). Ponadto określa się **tłumienie bezwzględne** $\sigma = \text{Re} s_{1,2} = -\omega_n \zeta$ (jest to stała czasowa obwiedni przebiegu oscylacyjnego) oraz **pulsację rezonansową** $\omega_r = \text{Im} s_{1,2} = \omega_n \sqrt{1 - \zeta^2}$. Przy $\zeta = 0$ obowiązuje równość $\omega_r = \omega_n$. Przy $\zeta > 0$ otrzymuje się drgania tłumione do zera (rys. 2.53C), przy $\zeta < 0$ — drgania narastające. Opis analityczny rozwiązania swobodnego ma postać $e^{-\omega_n \zeta t} \cos \omega_n \sqrt{1 - \zeta^2} t$. Na rysunku 2.53D pokazano rodzinę charakterystyk częstotliwościowych amplitudowych przy różnych wartościach ζ .

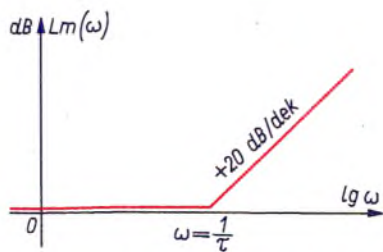
·2.54· Jaki wpływ na własności dynamiczne układu mają zera transmitancji?

W członach elementarnych omówionych wyżej licznik transmitancji był stały. W przypadku ogólniejszym licznik transmitancji jest także wielomianem zmiennej s , którego miejsca zerowe są **zerami** transmitancji. Po rozłożeniu wielomianów licznika i mianownika na czynniki pierwszego stopnia (lub ewentualnie drugiego stopnia „oscylacyjne”) transmitancja przybiera postać (dla przypadku czynników tylko 1 stopnia)

$$G(s) = \frac{(1 + s\tau_1) \dots (1 + s\tau_m)}{(1 + sT_1) \dots (1 + sT_n)}; \quad m \leq n$$

·2.53D· Rodzina charakterystyk częstotliwościowych członu oscylacyjnego przy $\zeta > 0$: a) charakterystyki amplitudowe; b) charakterystyki fazowe

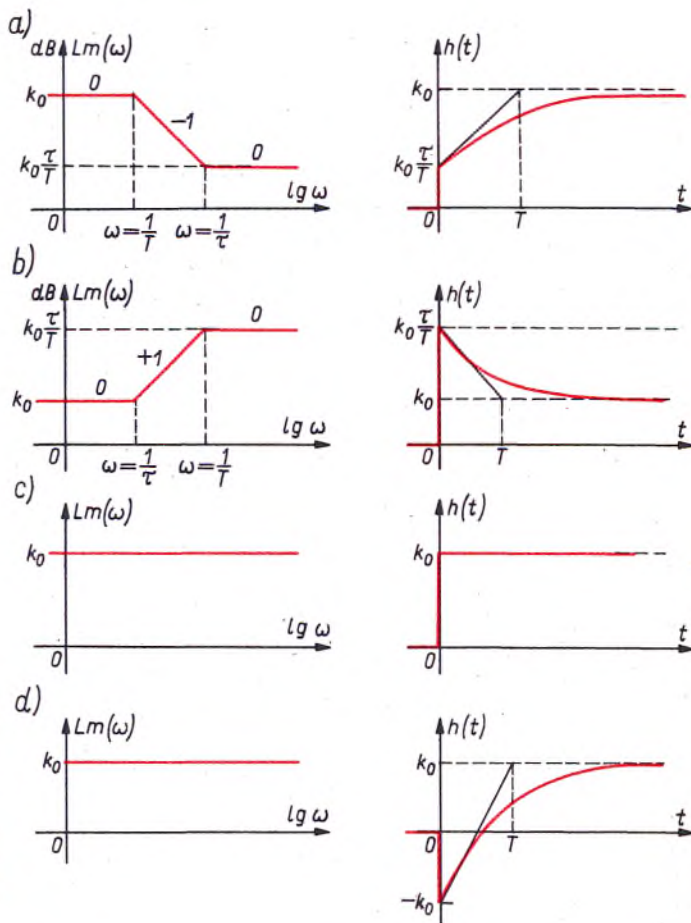




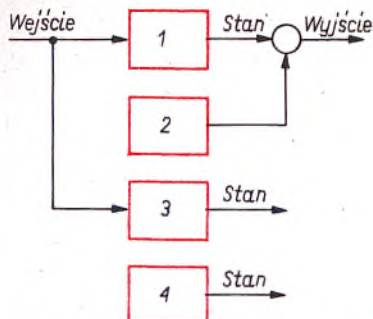
•2.54A• Wpływ czynnika $(1+s\tau)$ znajdującego się w liczniku transmitancji na przebieg charakterystyki częstotliwościowej układu

Czynniki licznika, typu $(1+s\tau)$, mają w pewnym sensie przeciwny wpływ na własności dynamiczne układu do podobnych czynników mianownika (rys. 2.54A). Człon $(1+s\tau)$ ma charakterystykę asymptotyczną o nachyleniach 0 dB/dek dla $\omega < \frac{1}{\tau}$, +20 dB/dek dla $\omega > \frac{1}{\tau}$ (gdyby występował sam czynnik $s\tau$, reprezentujący różniczkowanie idealne, to charakterystyka miałaby cały czas jedno nachylenie +20 dB/dek).

Na rysunku 2.54B pokazano cztery charakterystyczne przypadki elementarnego członu pierwszego rzędu — z jednym biegunem i jednym zerem transmitancji. Przypadki te różnią się tylko relacją stałych czasowych T i τ , jednak relacja ta ma bardzo istotny wpływ na właściwości układu, pomimo identyczności równania stanu. W szczególności przypadek (c) — skracanie się czynników licznika i mianownika — odpowiada niesterowalności układu, zaś przypadek (d) — nie-minimalnofazowości.

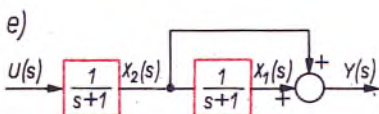
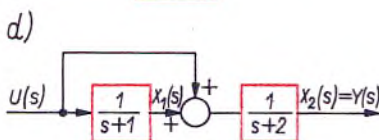
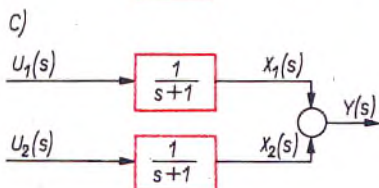
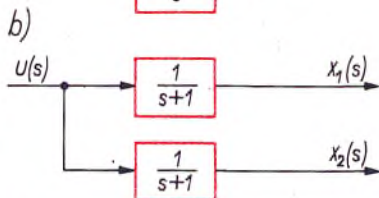
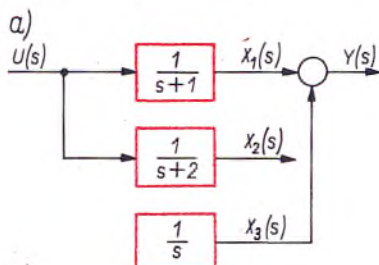


•2.54B• Charakterystyki częstotliwościowe asymptotyczne oraz charakterystyki skokowe członu o transmitancji $G(s) = k_0 \frac{1+s\tau}{1+sT}$, ($T > 0$): a) $T > \tau$ — człon proporcjonalno-inercyjny; b) $T < \tau$ — człon proporcjonalno-różniczkujący; c) $T = \tau$ — człon proporcjonalny (niesterowalność); d) $\tau = -T$ — człon nieminimalnofazowy



•2.55. Podział układu dynamicznego na części:

1 — część sterowalna i obserwowalna; 2 — część niesterowalna i obserwowalna; 3 — część sterowalna i nieobserwowalna; 4 — część niesterowalna i nieobserwowalna



•2.55. Kiedy występuje kompensacja bieguna przez zero transmitancji?

Kompensacja bieguna przez zero transmitancji, czyli skrócenie się czynników licznika i mianownika, występuje w przypadku, gdy układ jest **niesterowalny** albo **nieobserwowalny**. Wyjaśnienie tych zjawisk jest możliwe tylko na gruncie opisu układu w przestrzeni stanów. Opis transmitancyjny, w którym dokonano wszelkich możliwych skróceń wspólnych czynników licznika i mianownika, dotyczy tylko tej części układu, która jest sterowalna i obserwowalna. Ogólnie układ dynamiczny może zawierać obok części sterowalnej i obserwowalnej, część niesterowalną, ale obserwowalną, część nieobserwowalną, ale sterowalną, wreszcie część niesterowalną i nieobserwowalną, co można schematycznie przedstawić, tak jak na rys. 2.55. Cechą układu niesterowalnego lub nieobserwowalnego jest większa liczba wartości własnych niż biegunów transmitancji.

Jeśli kompensowany biegun i zero leżą w lewej płaszczyźnie zmiennej zespolonej, to można sobie wyobrazić sytuację kompensacji przybliżonej (równości przybliżonej bieguna i zera), natomiast w prawej półpłaszczyźnie kompensacja (przybliżona czy nawet idealna) nie ma sensu fizycznego, gdyż zawsze niestabilność spowodowana obecnością bieguna w prawej półpłaszczyźnie będzie zjawiskiem dominującym.

•2.56. Jaki jest sens pojęć sterowalności i obserwowalności?

Rozważmy układ swobodny $\dot{x} = Ax$, x — n -wymiarowy, z pojedynczą (skalarną) zmienną sterującą u . Powstaje pytanie, czy przy dowolnym warunku początkowym $x(0) = x^0$ można za pomocą tego sterowania u sprowadzić w **skończonym czasie** stan układu do wybranego punktu, np. $x = 0$. Jeśli tak, układ nazywa się **sterowalnym**. Istotą tej właściwości jest fakt, że każda ze zmiennych stanu układu sterowalnego powinna zależeć od sterowania w nieco inny sposób, tak aby sterowanie nie powodowało identycznych zmian poszczególnych współrzędnych stanu.

Podobnie rozważmy układ n -tego rzędu bez wymuszenia, ale z jednym wyjściem y . Jeśli dzięki obserwacji przebiegu w skończonym przedziale czasu (t_0, t) przy znanym wymuszeniu można określić wartości wszystkich zmiennych stanu w chwili t , to układ nazywa się **obserwowalnym**. Sens pojęcia sprowadza się do tego, czy poszczególne zmienne stanu wpływają na wyjście w sposób na tyle niezależny od siebie, że umożliwia to odróżnienie ich wpływu w zaobserwowanym przebiegu y .

Dla lepszego wyjaśnienia tych pojęć na rys. 2.56 pokazano kilka przykładów układów niesterowalnych lub nieobserwowalnych. Jak widać, efekty niesterowalności i nieobserwowalności wiążą się z pewną „niezręcznością” w sposobie wprowadzenia sterowania lub wyprowadzania wyjścia układu — w stosunku do układu swobodnego. Rysunek (a)

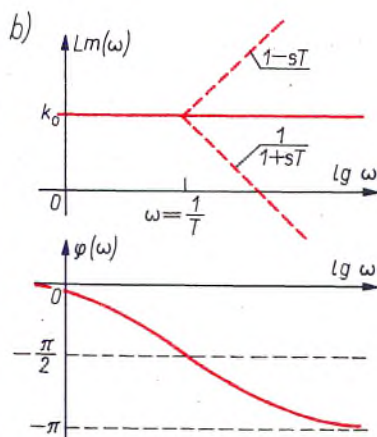
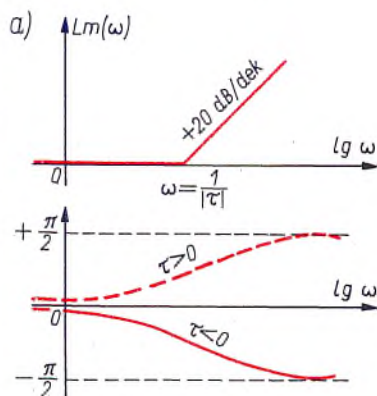
•2.56. Przykłady niesterowalności i nieobserwowalności: a) zmienna x_3 niesterowalna, zmienna x_2 nieobserwowalna; b) niesterowalność; c) nieobserwowalność; d) e) przypadki ze skracaniem się biegunów i zer transmitancji

obrazuje najbardziej jaskrawy przykład, kiedy strukturę układu swobodnego można przedstawić jako kilka niezależnych (równoległych) ogniw, a sterowanie lub wyjście są „odcięte” od niektórych zmiennych stanu. Nieco bardziej skomplikowana sytuacja ma miejsce na rys. (b) i (c), gdzie układ swobodny składa się z dwu identycznych ogniw równoległych; nie można jednak w przypadku (b) sterować niezależnie obydwu zmiennych stanu, a w przypadku (c) w przebiegu y nie można rozróżnić oddzielnego wpływu obydwu zmiennych stanu. Przypadki (d) i (e) są najbardziej mylące. Dopiero po przekształceniu schematów widać, że w obydwu przypadkach w transmitancji wypadkowej występuje skrócenie czynników $(s+2)$ w liczniku i mianowniku. W przypadku (d) oznacza to, że przebieg zmiennej x_2 wymuszony przez u pokrywa się z x_1 ; niezerowe warunki początkowe ujawniają jednak w przebiegu y obydwie zmienne x_1 i x_2 . W przypadku (e) jest odwrotnie: sterowanie obydwu zmiennych stanu jest możliwe bez przeszkód, natomiast na „niezręcznie” wybranym wyjściu nie jest widoczna zmienna x_2 , w tym również jej przebieg swobodny przy $u = 0$ i $x_2(0) \neq 0$. Tak więc skracanie się biegunów i zer transmitancji jest sygnałem, że w danym układzie występuje efekt niesterowalności albo nieobserwowalności. W przypadku (c) z rys. 2.54B podany układ jest niesterowalny — jego stanu nie można zmienić, z punktu widzenia opisu wejście-wyjście jest to układ bezinercyjny.

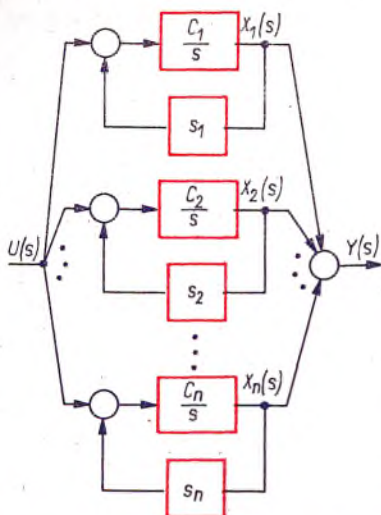
·2.57· Co to jest układ minimalnofazowy?

Jest to układ, w którym przebieg charakterystyki fazowej jest określony jednoznacznie przez charakterystykę amplitudową. Ścisły związek pomiędzy tymi charakterystykami wyraża tzw. całka Bodego. Istota tego związku sprowadza się do tego, że jeśli w szerokim przedziale ω charakterystyka logarytmiczna $Lm(\omega)$ ma nachylenie $k \cdot 20$ dB/dek, to charakterystyka fazowa w tym przedziale ustala się na poziomie $k \frac{\pi}{2}$. W szczególności dla $\omega \rightarrow \infty$ granica przesunięcia fazy

układu minimalnofazowego wynosi $-(n-m) \frac{\pi}{2}$, przy czym n — stopień mianownika, m — stopień licznika transmitancji (w przypadku gdy transmitancja jest ilorazem wielomianów zmiennej s). Układ nieminimalnofazowy wykazuje większe ujemne przesunięcie fazy niż określone wyżej — stąd nazwa. Nieminimalnofazowość może być spowodowana wystąpieniem zera lub bieguna transmitancji w prawej półpłaszczyźnie, ujemną wartością współczynnika wzmocnienia, a także może być efektem „rozłożoności” obiektu (patrz pyt. 2.72 do 2.81), w szczególności zawierającego opóźnienie (czynnik e^{-sT_0} jest nieminimalnofazowy). W wielu obiektach regulacji, zwłaszcza dużych obiektach „przemysłowych” nieminimalnofazowość jest cechą typową. Na rysunku 2.57 zilustrowano nieminimalnofazowość wnoszoną do u-



·2.57· Charakterystyki częstotliwościowe układów nieminimalnofazowych: a) nieminimalnofazowość spowodowana obecnością czynnika $(1+s\tau)$ w liczniku transmitancji (przy $\tau > 0$ i $\tau < 0$); b) charakterystyki członu o transmitancji $G(s) = k_0 \frac{1-sT}{1+sT}$



•2.58A• Struktura układu dynamicznego o macierzy diagonalnej A (transmitancji rozłożonej na ułamki proste)

kładu przez zero transmitancji leżące w prawej półpłaszczyźnie. Charakterystyka amplitudowa na rys. 2.57b jest identyczna z charakterystyką członu proporcjonalnego, natomiast faza zmienia się od 0 do $-\pi$, co dobitnie podkreśla znaczenie pojęcia nieminimalnofazowości.

•2.58• Jak należy wybierać zmienne stanu układu o danej transmitancji?

Jeśli jest dana transmitancja, to rząd układu należy utożsamiać ze stopniem wielomianu mianownika transmitancji. Spośród różnych możliwych sposobów wyboru zmiennych stanu kilka jest szczególnie ciekawych. Ze względów rachunkowych najbardziej pożądanym jest sposób, w którym otrzymuje się macierz A układu w postaci diagonalnej, przy czym na przekątnej tej macierzy występują poszczególne wartości własne (bieguny transmitancji) układu. Model taki można uzyskać rozkładając transmitancję na ułamki proste:

$$G(s) = \frac{b_m s^m + \dots + b_1 s + b_0}{s^n + \dots + a_1 s + a_0} = \frac{L(s)}{(s-s_1) \dots (s-s_n)} = \frac{C_1}{s-s_1} + \dots + \frac{C_n}{s-s_n}$$

Odpowiada to strukturze pokazanej na rys. 2.58A. Jako zmienne stanu wybiera się wtedy zmienne poszczególnych członów elementarnych połączonych równolegle i otrzymuje się opis w przestrzeni stanów o macierzach

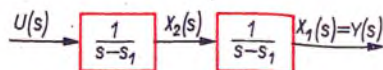
$$A = \begin{bmatrix} s_1 & 0 \\ & \ddots \\ 0 & s_n \end{bmatrix}; \quad b = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}; \quad c = [1 \dots 1]$$

Zasadnicze wady tego opisu wynikają z braku ogólności metody, ponieważ przy wielokrotnych wartościach własnych nie jest możliwy rozkład na ułamki o mianownikach tylko pierwszego stopnia. Na przykład układ o transmitancji $\frac{1}{(s-s_1)^2}$ nie daje się przedstawić jako struktura równoległa, lecz tylko jako szeregową, jak na rys. 2.58B, co odpowiada macierzy A o postaci tzw. klatki Jordana

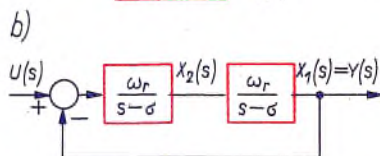
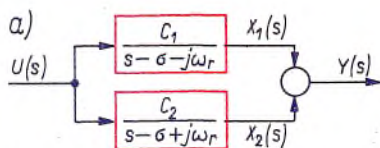
$$A = \begin{bmatrix} s_1 & 1 \\ 0 & s_1 \end{bmatrix}; \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}; \quad c = [1 \ 0]$$

W przypadku układów wyższych rzędów klatki Jordana mogą występować na przekątnej macierzy A . Innym brakiem metody równoległej jest niemożność uwzględnienia wartości własnych zespolonych, jeśli chcemy otrzymać model o współczynnikach rzeczywistych (ma to istotne znaczenie np. przy modelowaniu analogowych układów). Elementarny składnik oscylacyjny transmitancji ma postać (por. pyt. 2.53)

$$\frac{C}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{C}{(s - \sigma - j\omega_r)(s - \sigma + j\omega_r)}$$



•2.58B• Struktura układu o podwójnej wartości własnej



• 2.58C. Struktura elementarnego członu oscylacyjnego o transmitancji C

$\frac{s^2 + 2\zeta\omega_n s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$: a) z parametrami zespolonymi; b) z parametrami rzeczywistymi

Odpowiada ona strukturom pokazanym na rys. 2.58C, zaś macierz A można przedstawić albo w postaci zespolonej

$$A = \begin{bmatrix} \sigma + j\omega_r & 0 \\ 0 & \sigma - j\omega_r \end{bmatrix}$$

albo, co bardziej przydatne, w postaci rzeczywistej

$$A = \begin{bmatrix} \sigma & \omega_r \\ -\omega_r & \sigma \end{bmatrix}$$

A więc przy wartościach własnych zespolonych na przekątnej macierzy A układu wyższego rzędu mogą się również pojawić klatki o powyższej postaci.

Cechy ogólności (kanoniczności) ma sposób wyboru zmien-nych stanu jako tzw. **współrzędnych fazowych**, spełniających zależności

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= -a_0 x_1 - a_1 x_2 - \dots - a_{n-1} x_n \end{aligned}$$

przy czym współczynniki $a_0, a_1, \dots, a_{n-1}$ są identyczne ze współczynnikami wielomianu charakterystycznego (mianownika transmitancji) $s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0$. Macierz A , odpowiadająca powyższemu układowi równań stanu, ma postać tzw. **macierzy stowarzyszonej z wielomianem**

$$A = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ -a_0 & -a_1 & \dots & -a_{n-1} \end{bmatrix}$$

Jeśli przyjąć za punkt wyjścia transmitancję nie zawierającą zera, a więc złożoną tylko z wielomianu w mianowniku $M(s)$

$$\frac{Y(s)}{U(s)} = \frac{b_0}{s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0} = \frac{1}{M(s)}$$

to łatwo sprawdzić, że odpowiada to przyjęciu $y = b_0 x_1$ oraz wprowadzeniu wyrazu u do równania, z którego znajduje się x_n , a zatem macierzy (wektorów)

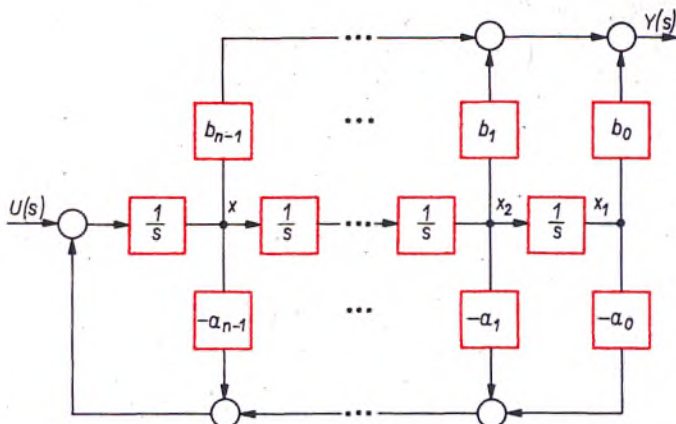
$$b = [0 \dots 0 \ 1]^T; \quad c = [b_0 \ 0 \dots 0]$$

Jeśli transmitancja zawiera zera, to można ją napisać jako

$$\begin{aligned} G(s) &= \frac{b_m s^m + \dots b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \dots + a_1 s + a_0} = \\ &= \frac{b_m s^m}{M(s)} + \dots + \frac{b_1 s}{M(s)} + \frac{b_0}{M(s)} \end{aligned}$$

Wyjście y składa się teraz z sumy kolejnych pochodnych

2.58D• Struktura kanoniczna sterowalna układu dynamicznego dla $m = n-1$



poprzednio użytej zmiennej x_1 , a więc kolejnych współrzędnych fazowych, ze współczynnikami $b_0, \dots, b_m$

$$y = b_0 x_1 + b_1 x_2 + \dots + b_m x_{m+1}$$

Jest to możliwe przy $m < n$ (a więc po wyłączeniu z transmisji ewentualnego składnika stałego d). W efekcie otrzymujemy n -wymiarowy wektor $c = [b_0 \ b_1 \ \dots \ b_m \ 0 \ \dots \ 0]$, natomiast wektor b jest identyczny jak poprzednio. Przedstawiony model odpowiada zawsze układowi sterowalnemu. Strukturę modelu pokazano na rys. 2.58D.

Podobną strukturę ma model, w którym przyjmujemy $y = x_n$, zmienne stanu ponumerujemy odwrotnie i uzależnimy wszystkie równania stanu od x_n . Równania układu o transmitancji z zerami mają wtedy postać

$$\dot{x}_1 = -a_0 x_n + b_0 u$$

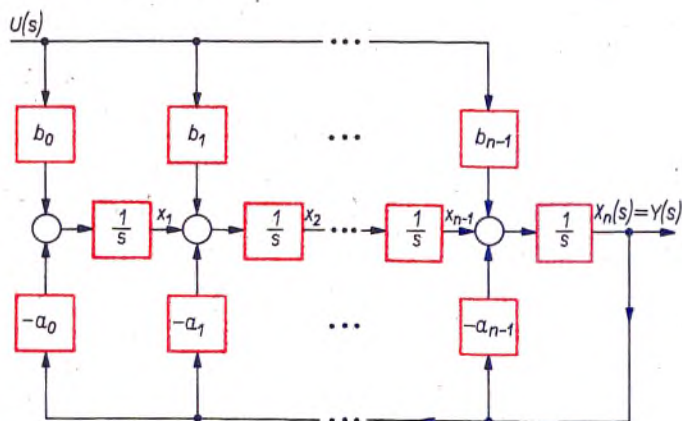
$$\dot{x}_2 = x_1 - a_1 x_n + b_1 u$$

$$\vdots$$

$$\dot{x}_{m+1} = x_m - a_m x_n + b_m u$$

$$\vdots$$

$$\dot{x}_n = x_{n-1} - a_{n-1} x_n$$



2.58E• Struktura kanoniczna obserwowalna układu dynamicznego dla $m = n-1$

Postać tych równań może wydawać się niejasna, jednak po narysowaniu schematu widać, że jest to struktura systematyczna i ogólna (rys. 2.58E). Przy $b_1 = b_2 = \dots = b_m = 0$ otrzymuje się oczywiście model transmitancji bez zer. Warto zwrócić uwagę na pewną symetrię macierzy występujących w tym opisie w stosunku do poprzedniego przypadku:

$$A = \begin{bmatrix} 0 & & & & -a_0 \\ 1 & & 0 & & -a_1 \\ & \ddots & & & \vdots \\ & & 1 & & -a_m \\ & & & \ddots & \vdots \\ & & & & 1 & -a_{n-1} \end{bmatrix}; \quad b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \\ 0 \\ \vdots \\ 0 \end{bmatrix}; \quad c = [0 \dots 0 \ 1]$$

Model ten odpowiada zawsze układowi obserwowalnemu.

•2.59. Które z omówionych zależności obowiązują w przypadku układów niestacjonarnych?

W przypadku układu niestacjonarnego (tj. układu liniowego, ale o parametrach zależnych od czasu) podstawową metodą opisu pozostają równania różniczkowe (równania stanu). Postać tych równań nie ulega zasadniczej zmianie — równania układu o r wejściach, m wyjściach i n zmiennych stanu można zapisać jako

$$\dot{x} = A(t)x + B(t)u$$

$$y = C(t)x + D(t)u$$

Jednak przy takiej postaci nie można stosować przekształceń upraszczających analizę ze względu na zależność od czasu wszystkich macierzy współczynników. Nie można podać metody ogólnej rozwiązywania takich równań, nie można zdefiniować wartości własnych układu i nie można stosować przekształceń Laplace'a. Określenia podstawowych własności układu, takich jak stabilność, sterowalność, obserwowalność — stają się nadzwyczaj trudne.

•2.60. Czy dla układu niestacjonarnego można określić macierz podstawową rozwiązań?

Tak, można zdefiniować macierz podstawową, jednak obliczenie tej macierzy jest najczęściej bardzo trudne. Jeśli rozważyć układ bez wymuszeń, to równanie stanu dla przypadku skalarnego

$$\dot{x} = a(t)x(t); \quad x(0) = x^0$$

ma rozwiązanie dość proste

$$x(t) = x^0 \exp \int_0^t a(\tau) d\tau$$

Jednakże, niestety, **nie można** przez analogię przyjąć macierzy podstawowej w postaci $\exp \int_0^t A(\tau) d\tau$, chyba że macierz $A(\tau)$ jest diagonalna (albo stała, ale wtedy mamy do czynienia z przypadkiem stacjonarnym).

Można ogólnie wprowadzić macierz podstawową rozwiązani, oznaczając ją $\Phi(t, t_0)$, przy czym istotna jest zależność tej macierzy od dwu argumentów: t — czasu bieżącego i t_0 — chwili początkowej (niekoniecznie zerowej). Macierz ta powinna spełniać zależności:

$$x(t) = \Phi(t, t_0)x(t_0)$$

$$\Phi(t_0, t_0) = I$$

$$\frac{d}{dt}\Phi(t, t_0) = A(t)\Phi(t, t_0)$$

$$\Phi(t, t_0) = \Phi(t, t_1)\Phi(t_1, t_0)$$

$$\Phi^{-1}(t, t_0) = \Phi(t_0, t)$$

Efektywne obliczenie macierzy $\Phi(t, t_0)$ na podstawie macierzy $A(t)$ jest zawsze bardzo trudne.

·2.61· Czy dla układu niestacjonarnego można określić transmitancję?

Nie, w żadnym przypadku. Nie można także oczywiście podać żadnych parametrów typu stałych czasowych, współczynników tłumienia, czy częstotliwości rezonansowych. W prostszych przypadkach można niekiedy określić współczynnik wzmocnienia zależny od czasu — jeśli w układzie można wyodrębnić część niestacjonarną statyczną i część dynamiczną stacjonarną.

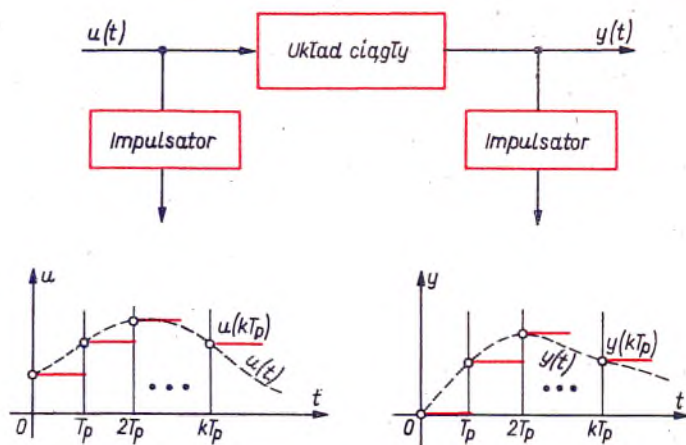
·2.62· Czy istnieje możliwość sformułowania opisu typu wejście-wyjście dla układu niestacjonarnego?

Tak, za pomocą operatora podobnego do całki spłotowej. Wynika to stąd, że składową wymuszoną rozwiązania dla stanu x można przedstawić za pomocą macierzy podstawowej $\Phi(t, t_0)$ jako

$$x(t) = \int_{t_0}^t \Phi(t, \tau)B(\tau)u(\tau)d\tau$$

zaś przebieg $y(t)$ wynika z równania wyjścia, $y(t) = C(t)x(t) + D(t)u(t)$. Powyższa całka przypomina całkę spłotową układu stacjonarnego, jednakże „funkcja wagi” (a raczej jądro operatora całkowego) $\Phi(t, \tau)B(\tau)$ nie ma teraz interpretacji macierzy charakterystyk impulsowych i zależy od dwu argumentów t i τ (zmiennych całkowania), a nie od ich różnicy ($t - \tau$) — jak w przypadku stacjonarnym. Tak więc możliwość sformułowania opisu typu wejście-wyjście istnieje, jednakże efektywne obliczenia związane z tym opisem są bardzo trudne.

·2.63· Ilustracja idei dyskretnego opisu układu ciągłego



·2.63· Kiedy stosuje się modele dyskretnie układów dynamicznych?

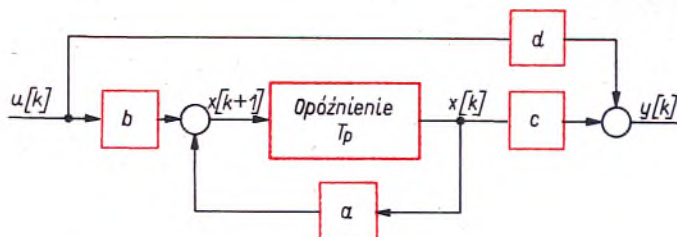
Potrzeba stosowania modeli dyskretnych wynika ze względów technicznych (zwłaszcza pomiarowych) i obliczeniowych. W zasadzie każdy układ rozważany makroskopowo należałoby traktować jako ciągły w czasie. Jednak w wielu przypadkach sygnały z natury ciągle poddaje się **dyskretyzacji**, po czym dopiero następuje dalsze przetwarzanie tych sygnałów. Dyskretyzacja (inaczej próbkowanie, impulsowanie) polega na pobieraniu — najczęściej okresowo — próbek wartości, a więc zamiast sygnału ciągłego $x(t)$ wytwarza się ciąg $x(kT_p)$, przy czym T_p oznacza okres próbkowania. Zazwyczaj urządzenia próbkujące dokonują ponadto zapamiętania (zatrzymania) wartości $x(kT_p)$ aż do następnej chwili próbkowania — w efekcie powstaje tzw. sygnał schodkowy (czyli sygnał z ekstrapolacją rzędu zerowego). Można to schematycznie przedstawić w sposób pokazany na rys. 2.63, gdzie zaznaczono, że analizie podlega pewien obiekt ciągły, którego sygnały wejściowy u i wyjściowy y podlegają okresowemu próbkowaniu i zapamiętywaniu. Informacja o sygnałach u i y w postaci dyskretniej stanowi podstawę tworzenia dyskretnych modeli układów. Sygnały o tej postaci nadają się zwykle bezpośrednio do przetwarzania cyfrowego, co jest główną przyczyną zajmowania się modelami dyskretnymi.

·2.64· Czym różni się podstawowa metoda opisu układów dyskretnych od opisu układów ciągłych?

Ponieważ w układach dyskretnych czas jest zmienną nieciągłą, więc podstawowe równanie stanu układu ma postać równania **różnicowego**, a nie różniczkowego. W przypadku najprostszym (pierwszego rzędu) równanie stanu jest analogiczne do równania z pyt. 2.34

$$x[k+1] = ax[k] + bu[k]$$

•2.64• Struktura równań najprostszego układu dyskretnego



Symbolem $x[k]$ oznaczać będziemy wartości w dyskretnych punktach czasu, tzn. $x[k] = x(kT_p)$. Na rysunku 2.64 pokazano schematycznie strukturę tego równania wraz z równaniem wyjścia. Widać, że jest to struktura zupełnie analogiczna do przypadku ciągłego, z tym że zamiast integratora występuje **element opóźniający** o jeden okres T_p .

Układ rzędu n ma n zmiennych stanu, wobec czego przy sterowaniu skalarnym i wyjściu skalarnym otrzymuje się w zapisie wektorowo-macierzowym równania

$$\begin{aligned}x[k+1] &= Ax[k] + bu[k] \\ y[k] &= cx[k] + du[k]\end{aligned}$$

przy czym x jest wektorem n -wymiarowym, A — macierzą współczynników $n \times n$, b i c — n -wymiarowymi wektorami współczynników.

W przypadku układu dyskretnego o r wejściach i m wyjściach otrzymuje się analogicznie do przypadku ciągłego

$$\begin{aligned}x[k+1] &= Ax[k] + Bu[k] \\ y[k] &= Cx[k] + Du[k]\end{aligned}$$

Struktura modelu układu dyskretnego liniowego różni się więc od układu ciągłego tylko zastąpieniem integratorów elementami opóźniającymi o jeden okres T_p . Opis ten ma identyczne właściwości formalne jak opis układu ciągłego, w szczególności — wybór zmiennych stanu nie jest jednoznaczny.

•2.65• Jaką postać ma opis typu wejście-wyjście dla układów dyskretnych?

Równanie różnicowe typu wejście-wyjście można zapisać

$$\begin{aligned}y[k+n] + a_{n-1}y[k+n-1] + \dots + a_1y[k+1] + a_0y[k] &= \\ = b_nu[k+m] + \dots + b_1u[k+1] + b_0u[k]\end{aligned}$$

Zastosowanie przekształcenia $\mathcal{Z}$ sprowadza to równanie do równania algebraicznego (transformata $\mathcal{Z}$ funkcji przesuniętej, przy zerowych warunkach początkowych, jest równa transformacie funkcji nie przesuniętej pomnożonej przez z)

$$\begin{aligned}(z^n + a_{n-1}z^{n-1} + \dots + a_1z + a_0)Y(z) &= \\ = (b_nz^m + \dots + b_1z + b_0)U(z)\end{aligned}$$

Na podstawie tego równania określamy **transmitancję dyskretną**

$$G(z) = \frac{Y(z)}{U(z)} = \frac{b_m z^m + \dots + b_1 z + b_0}{z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0}$$

Tak jak w przypadku układu ciągłego, znając transmitancję układu dyskretnego można określić odpowiedź wymuszoną w dziedzinie transformat

$$Y(z) = G(z) U(z)$$

przy wymuszeniu u o transformacie $U(z)$.

Odpowiednikiem całki spłotowej jest „suma spłotowa”

$$y[k] = \sum_{i=0}^{k-1} g[k-i] u[i]$$

przy czym $g[k-i]$ jest dyskretną charakterystyką impulsową, stanowiącą transformatę odwrotną transmitancji $G(z)$.

·2.66· Jakie są zależności między opisem typu wejście-wyjście a opisem w przestrzeni stanów dla układów dyskretnych?

Zależności te wykazują niemal zupełną analogię do przypadku układu ciągłego. Przekształcając do dziedziny zmiennej z równania stanu i wyjścia układu dyskretnego (pyt. 2.64), otrzymuje się przy zerowych warunkach początkowych

$$zX(z) = AX(z) + BU(z)$$

$$Y(z) = CX(z) + DU(z)$$

Po identycznych przekształceniach algebraicznych, jak w przypadku ciągłym, uzyskuje się wyrażenie określające transmitancję macierzową dyskretną

$$G(z) = C(zI - A)^{-1}B + D$$

Stąd wynika natychmiast wniosek, że podobnie jak dla układu ciągłego — pierwiastki równania charakterystycznego o postaci $\det(zI - A) = 0$, czyli wartości własne macierzy A , są wartościami własnymi układu i biegunami transmitancji dyskretnych. Wartości własne z_i ($i = 1, \dots, n$) determinują pojawianie się w przebiegach rozwiązań członów typu z_i^k lub w przypadku wartości własnych wielokrotnych — typu kz_i^k , $k^2z_i^k$ itd. Macierz $(zI - A)^{-1}$ pojawia się także w transformacie rozwiązywania swobodnego, bowiem dla

$$x[k+1] = Ax[k]; \quad x[0] = x^0$$

otrzymuje się

$$zX(z) - zx^0 = AX(z)$$

czyli

$$X(z) = (zI - A)^{-1}zx^0$$

Ponieważ

$$x[k] = \mathcal{Z}^{-1} \{ (zI - A)^{-1} z x^0 \}$$

więc transformata odwrotna wyrażenia $(zI - A)^{-1}z$ jest to **macierz podstawowa** rozwiązań równania stanu. Jednak dla układu dyskretnego tę macierz podstawową można wyrazić znacznie prościej, korzystając z metody rekurencyjnej rozwiązywania równania stanu (podstawiając kolejno $k = 0, 1, 2, \dots$):

$$\begin{aligned}x[1] &= Ax^0 \\x[2] &= A^2x^0 \\&\vdots \\x[k] &= A^kx^0\end{aligned}$$

A więc macierz podstawowa jest równa A^k . Równanie wyjścia układu dyskretnego wyrażone za pomocą macierzy podstawowej ma postać

$$y[k] = C\{A^kx^0 + \sum_{i=0}^{k-1} A^{k-i-1}Bu[i]\} + Du[k]$$

W wyrażeniu tym można łatwo rozpoznać składową swobodną (przy niezerowych warunkach początkowych) i wymuszoną (w postaci sumy spłotowej uzyskanej zupełnie podobnie jak w przypadku układu ciągłego).

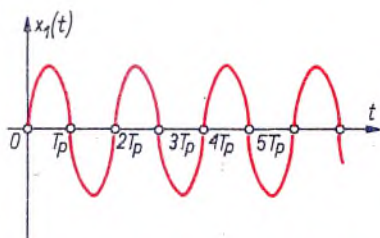
Analogia do przypadku ciągłego obejmuje również przejście odwrotne — od opisu transmitancyjnego do opisu we współrzędnych stanu. Jako zmienne stanu najprościej jest wybrać n wartości przesuniętych $y[k], y[k+1], \dots, y[k+n-1]$. Budowa modelu stanowego układu dyskretnego podlega identycznym zasadom jak omówione w pyt. 2.58 dla układu ciągłego, z tym że w schematach strukturalnych należy zastąpić integratory członami opóźniającymi o jeden okres T_p .

·2.67· Jaki jest podstawowy warunek stabilności układu dyskretnego?

Ponieważ elementarny składnik rozwiązania równania różnicowego jest typu z_i^k (z_i — wartość własna), więc warunkiem stabilności jest, aby wartości bezwzględne wszystkich wartości własnych były mniejsze od jedności:

$$|z_i| < 1, \quad i = 1, \dots, n$$

Przy $|z_i| = 1$ obowiązują podobne zastrzeżenia, jak przy $\text{Re } s_i = 0$ w przypadku ciągłym — tzn. przy wartości własnej rzeczywistej jednokrotnej lub parze zespolonej sprzężonej warunek $|z_i| = 1$ zapewnia stabilność, lecz nie zapewnia stabilności asymptotycznej. W pozostałych przypadkach układ jest niestabilny. Zwróćmy przy okazji uwagę, że pojęcie „oscylacyjności” przebiegu dyskretnego niekoniecznie wiąże się z zespolonymi wartościami własnymi. Przy rzeczywistej, ujemnej wartości własnej przebieg wykazuje przemienność znaku co okres.



•2.68• Zjawisko ukrytych oscylacji, niewidocznych w przebiegu dyskretnym

•2.68• Czy pojęcia sterowalności i obserwowalności odnoszą się także do układów dyskretnych?

Tak, sens tych pojęć jest w zasadzie identyczny jak w przypadku ciągłym (pyt. 2.56). Należy podkreślić, że opis dyskretny pozwala na wyraźniejszą interpretację możliwości sprowadzenia stanu układu do punktu $x = 0$ (sterowalność) lub określenia wektora stanu x na podstawie wyjścia y (obserwowalność) w **skończonym czasie**, ponieważ warunki skończonego czasu sterowania lub obserwacji oznaczają po prostu **skończoną liczbę** okresów impulsowania czyli skończoną liczbę wartości dyskretnych $u(kT_p)$ lub $y(kT_p)$. Okazuje się przy tym, że liczba tych wartości dyskretnych powinna być równa rzędowi układu dyskretnego. Warunki sterowalności i obserwowalności dla układu dyskretnego podlegają jednak dodatkowemu ograniczeniu: okres impulsowania T_p nie może stanowić wielokrotności ewentualnego okresu oscylacji występujących w części ciągłej układu opisywanego za pomocą modelu dyskretnego — aby nie wystąpiło zjawisko tzw. ukrytych oscylacji, zilustrowane na rys. 2.68.

•2.69• Czy ten sam obiekt rzeczywisty można opisywać za pomocą modelu ciągłego i dyskretnego?

Tak, istnieją przy tym ściśle związki pomiędzy jedną a drugą metodą opisu. Jeśli za punkt wyjścia uznać równania stanu modelu ciągłego, to w pierwszym przybliżeniu równania różnicowe modelu dyskretnego uzyskuje się po prostu zastępując pochodne ilorazami różnicowymi

$$\frac{dx}{dt} \approx \frac{x(t+T_p) - x(t)}{T_p} = \frac{x[k+1] - x[k]}{T_p}$$

Przybliżenie takie jest tym lepsze, im mniejsza jest wartość T_p ; normalizacja skali czasu względem T_p umożliwia przyjęcie $T_p = 1$.

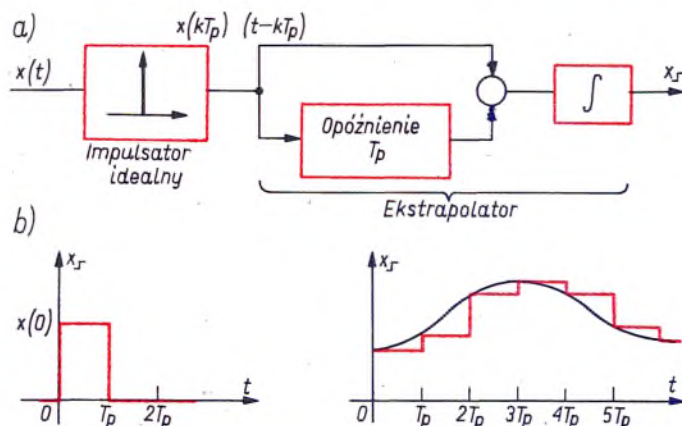
Również opis transmitancyjny modelu ciągłego i dyskretnego można wzajemnie uzależnić. Ponieważ charakterystyka impulsowa dyskretna jest to zdyskretyzowany przebieg charakterystyki impulsowej ciągłej, a odpowiednie transmitancje — dyskretna i ciągła — są transformatami ($\mathcal{Z}$ lub $\mathcal{L}$) charakterystyk impulsowych, stąd wniosek, że transmitancję dyskretną obiektu ciągłego można wyznaczyć na podstawie transmitancji ciągłej i odwrotnie. Zabieg taki nazywa się **przekształcaniem transformat** i jest opisany ściśłymi zależnościami, zauważmy jednak, że do celów praktycznych wystarczy zestawić obok siebie pary transformat ciągłych i dyskretnych tych samych funkcji czasu, aby w łatwy sposób przechodzić od jednego rodzaju opisu do drugiego. Ma to istotne znaczenie, ponieważ często zachodzi sytuacja, że układ rzeczywisty ma częściowo naturę ciągłą, częściowo dyskretną (np. obiekt sterowany ciągły, zaś regulator — dyskretny). W takiej sytuacji można opisywać obiekt ciągły

za pomocą zwykłej transmitancji, zwłaszcza jeśli składa się on z większej liczby członów ciągłych połączonych w różny sposób. Podobnie regulator impulsowy można opisywać jako układ dyskretny. Rozważania wymagają tylko punkty styku części ciągłej układu z częścią dyskretną. Przejście z sygnału ciągłego na dyskretny wymaga operacji próbkowania, której sens jest zrozumiały, a realizacja techniczna dość prosta. Przetwarzanie danych o postaci dyskretniej jest domeną urządzeń cyfrowych. Przejście z sygnału dyskretnego na ciągły („analogowy”) wymaga ekstrapolacji wartości w dyskretnych chwilach czasu, tak aby sygnał wprowadzany na obiekt ciągły miał postać ciągłą. W analizie złożonego układu impulsowego, zawierającego zarówno elementy o działaniu ciągłym jak i elementy faktycznie dyskretnie, należy uwzględnić — jako oddzielne elementy — **impulsatory i ekstrapolatory**, najczęściej zaś ich połączenie.

•2.70• Jaki jest najprostszy impulsator i ekstrapolator?

Pojęciowo elementarnym impulsatorem jest impulsator idealny, wytwarzający z sygnału ciągłego ciąg impulsów Diraca o wagach odpowiadających wartościom dyskretnym, tzn. jeśli sygnał ciągły $x(t)$ ma wartości dyskretnie $x(kT_p)$, to sygnał z impulsatora idealnego ma postać $x(kT_p)\delta(t - kT_p)$.

Realny układ impulsowy oczywiście nie wytwarza takich impulsów, natomiast jeśli za hipotetycznym impulsatorem idealnym wyobrazimy sobie człon formujący impulsy o „realnej” postaci, to taka para impulsator-człon formujący stanowi dobry model impulsatora rzeczywistego. Najczęściej działanie impulsatora rzeczywistego można przedstawić jako ekstrapolację rzędu zerowego, tzn. wytwarzanie przebiegu schodkowego x_r . W tym przypadku (rys. 2.70) można uznać, że przebieg ciągły przechodzi przez impulsator idealny, a następnie przez człon wytwarzający w odpowiedzi na impuls $\delta(t)$ — impuls prostokątny o szerokości T_p : $1(t) - 1(t - T_p)$. Wobec tego ekstrapolator ma transmitancję



•2.70• Wytwarzanie funkcji schodkowej: a) schemat struktury układu z impulsatorem idealnym i ekstrapolatorem; b) odpowiedź ekstrapolatora na pojedynczy impuls $x(0)$; c) funkcja schodkowa

(ciągłą) równą $\frac{1}{s} (1 - e^{-sT_F})$. Transmittancja dyskretna takiego ekstrapolatora jest tożsamościowo równa 1, ponieważ przebiegi ciągły i schodkowy nie różnią się w chwilach impulsowania. Natomiast dołączenie ekstrapolatora do obiektu ciągłego zmienia nieco łączną transmittancję obiektu „widzianą” od strony części impulsowej. W najprostszych przypadkach uwzględnienie ekstrapolatora wprowadza opóźnienie o jeden okres impulsowania.

2.71. Czy cecha niestacjonarności jest łatwa do uwzględnienia w przypadku układu dyskretnego?

Niestacjonarność układów dyskretnych powoduje podobne ograniczenia jak w przypadku ciągłym. Tak więc nie można operować transmittancjami, nie można określić wartości własnych itd. Rozwiązanie równań różnicowych układu niestacjonarnego jest jednak stosunkowo proste. Jeśli rozważymy równanie różnicowe stanu układu swobodnego (niestacjonarnego)

$$x[k+1] = A[k]x[k]; \quad x[0] = x^0$$

to stwierdzamy od razu, że rekurencyjna metoda rozwiązywania jest nadal słuszna, bowiem

$$x[1] = A[0]x^0$$

$$x[2] = A[1]x[1] = A[1]A[0]x^0$$

⋮

$$x[k] = A[k-1] \dots A[1]A[0]x^0$$

Stąd można określić macierz podstawową rozwiązań w postaci ogólnej

$$\Phi[k, k_0] = \prod_{i=k_0}^{k-1} A[i]$$

Macierz ta spełnia analogiczne postulaty jak w przypadku ciągłym, w szczególności

$$x[k] = \Phi[k, k_0]x[k_0]$$

Dość łatwo można również otrzymać składową wymuszoną rozwiązania równania stanu przy zerowym warunku początkowym. Jeśli zapiszemy równanie stanu jako

$$x[k+1] = A[k]x[k] + B[k]u[k]; \quad x[0] = 0$$

to

$$x[1] = B[0]u[0]$$

$$x[2] = A[1]x[1] + B[1]u[1] = A[1]B[0]u[0] + B[1]u[1]$$

$$x[3] = A[2]A[1]B[0]u[0] + A[2]B[1]u[1] + B[2]u[2]$$

⋮

Wprowadzając macierz podstawową Φ i zauważając, że

$$\Phi[k, k_0]\Phi^{-1}[i, k_0] = A^{k-1} \cdot \dots \cdot A^i, \text{ można napisać}$$

$$\mathbf{x}[k] = \Phi[k, k_0] \sum_{i=k_0+1}^k \Phi^{-1}[i, k_0] \mathbf{B}[i-1] \mathbf{u}[i-1]$$

Oczywiście pełne rozwiązanie składa się ze składowej swobodnej, zależnej od warunku początkowego $\mathbf{x}^0$, i składowej wymuszonej.

Względna prostota obliczeniowa wyrażeń (tylko obliczanie iloczynów macierzy) jest jednym z głównych argumentów przemawiających za potrzebą dyskretyzacji równań układu (zwłaszcza niestacjonarnego) w celu umożliwienia obliczeń numerycznych przy użyciu elektronicznej techniki obliczeniowej.

·2.72· Jaki jest podstawowy sposób opisu układów o parametrach rozłożonych?

Podstawowym sposobem opisu tych układów są równania różniczkowe o pochodnych cząstkowych. Wynika to stąd, że poza zmienną niezależną t (czasem) należy uwzględnić zależność od współrzędnych przestrzennych. W ogólnym przypadku występują trzy współrzędne przestrzenne, ale efekt rozłożoności parametrów można wyraźnie pokazać już przy uwzględnieniu jednego kierunku przestrzennego. Zajmiemy się tym ostatnim przypadkiem, wybraną zmienną przestrzenną (odległość) oznaczmy przez z . Jak już podano wcześniej w pyt. 2.14, stan układu o parametrach rozłożonych nawet w najprostszym przypadku nie daje się przedstawić jako wektor o skończonej liczbie składowych, ponieważ przy jednej interesującej nas zmiennej x trzeba tę zmienną przedstawić jako funkcję argumentów t i z , tzn. $x = x(t, z)$. W ustalonej chwili $t = t_0$ mamy do czynienia z **rozkładem** $x(t_0 = \text{const}, z)$ wzdłuż współrzędnej przestrzennej z , natomiast w ustalonych punktach $z = z_0$ otrzymujemy x jako przebieg będący funkcją tylko czasu $x(t, z_0 = \text{const})$.

·2.73· W jaki sposób można sterować układ o parametrach rozłożonych?

W rozwiązaniach równań o pochodnych cząstkowych uwzględnia się, obok warunków **początkowych** $x(t = t_0, z)$, **warunki brzegowe** $x(t, z = z_0)$. Warunek początkowy w chwili $t = t_0$ może być na ogół rozłożony na całej długości z . Sformułowanie warunków brzegowych w sposób dość istotny zależy od geometrii układu, w szczególności przy jednej zmiennej przestrzennej z — od tego, czy układ ma skończoną „długość” (np. $0 \leq z \leq Z$), czy też jest nieograniczony (np. $0 \leq z < \infty$). Tak więc na przebieg rozwiązania można oddziaływać za pomocą czynnika (sterowania) **rozłożonego** wzdłuż całego układu lub też **skupionego** tylko na jego brzegach. Większość przypadków praktycznych można w sposób zadowalający opisać rozważając tylko sterowanie skupione, ponieważ w układach o parametrach rozłożonych, spotykanych w technice, zazwyczaj można jednoznacznie określić punkt będący „początkiem” układu lub jego „końcem”.

·2.74· W jaki sposób obserwuje się stan układu o parametrach rozłożonych?

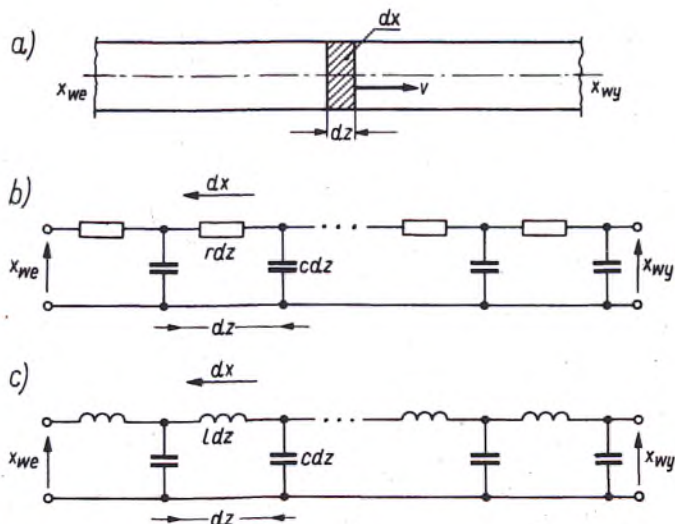
Ścisła obserwacja wymagałaby pomiaru rozkładów zmiennej x względem zmiennej przestrzennej z w każdej chwili czasu t . Jest to oczywiście wymaganie nierealne. Praktycznie jesteśmy w stanie mierzyć przebiegi czasowe interesującej nas zmiennej x w jednym lub kilku odosobnionych punktach osi z . Sytuacja taka odpowiada wyjściu **skupionemu** układu.

·2.75· Czy układ o wejściu i wyjściu skupionym można opisać podobnie jak układ o parametrach skupionych?

Tak, rozważanie skupionego wejścia i skupionego wyjścia umożliwia traktowanie sygnałów w tych punktach w sposób analogiczny do układu o parametrach skupionych. W szczególności, dla układów liniowych i stacjonarnych, można określić transmitancję wejście-wyjście lub charakterystykę impulsową, dzięki której można sformułować opis za pomocą całki spłotowej. Trzeba jednak pamiętać, że transmitancja wejście-wyjście wyjaśnia zachowanie się układu tylko przy zerowych warunkach początkowych i brzegowych; transmitancja taka nie odzwierciedla wpływu rozkładu początkowego $x(t_0, z)$ ani sterowania (czy tym bardziej zakłócenia) rozłożonego.

·2.76· Jakie można podać przykłady typowych układów o parametrach rozłożonych?

Najczęściej spotykanymi układami o parametrach rozłożonych są układy, w których zachodzą zjawiska **przepływu**, **dyfuzji** lub zjawiska **falowe**. Dość rzadko występuje tylko jedno z wymienionych zjawisk. Tym niemniej zapoznanie się z modelami tych zjawisk daje pogląd o możliwych właściwościach układów o parametrach rozłożonych.



·2.76· Elementarne układy o parametrach rozłożonych: a) przepływ płynu; b) kontinuum RC; c) kontinuum LC

Zjawiska przepływu występują w układach z przepływem płynów (cieczy lub gazów) przez rurociągi lub z transportem masy za pomocą taśmociągów itp. — rys. 2.76a. Podstawowym parametrem jest **prędkość przepływu i opóźnienie przenoszenia zaburzenia** z wejścia na wyjście układu.

Zjawiska typu dyfuzji charakteryzują nie tylko właściwą dyfuzję, czyli wymianę masy (mieszanie), lecz także przewodzenie ciepła. Model dyfuzji odnosi się również do tzw. kontinuum RC (rys. 2.76b), tzn. linii elektrycznej o rezystancji i pojemności rozłożonych wzdłuż długości linii (a więc określonych na jednostkę długości). Idealne zjawisko dyfuzji nie wykazuje opóźnienia, tzn. wprowadzenie zaburzenia na wejściu powoduje, że odpowiedź na wyjściu pojawia się od razu i stopniowo narasta.

Zjawiska falowe występują w mechanice (akustyka, hydrodynamika) oraz w elektrotechnice (rozchodzenie się fal elektromagnetycznych, przepływ prądu przez tzw. linie długie). Prostim modelem układu o tego typu właściwościach jest kontinuum LC (rys. 2.76c), czyli linia elektryczna o indukcyjności i pojemności rozłożonych na długości.

·2.77· Jak opisuje się zjawisko przepływu?

Podstawowym modelem zjawiska przepływu jest równanie

$$\frac{\partial x(t, z)}{\partial t} = -v \frac{\partial x(t, z)}{\partial z}$$

wynikające wprost z bilansu przepływającego materiału (x reprezentuje ilość materiału, z jest odległością, parametr v określa prędkość przepływu). Poddając to równanie przekształceniu Laplace'a, w odniesieniu do zależności czasowych, otrzymuje się przy zerowym warunku początkowym dla zmiennej x równanie różniczkowe zwyczajne względem zmiennej z

$$v \frac{dX(s, z)}{dz} + sX(s, z) = 0$$

Równanie to ma oczywiste rozwiązanie

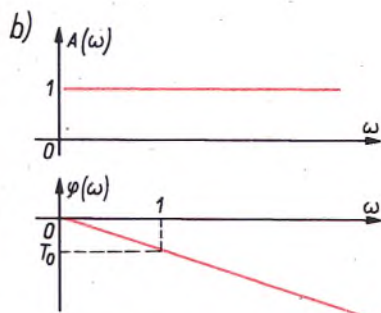
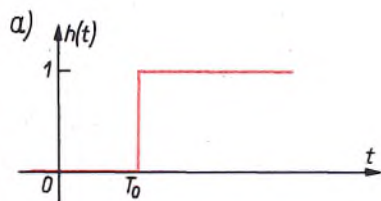
$$X(s, z) = e^{-\frac{s}{v}z} X(s, z)|_{z=0}$$

A więc, jeśli badamy transformaty $X(s, z)$ w punkcie $z = 0$ (wejście) i w pewnym punkcie „bieżącym” z , traktowanym jako wyjście, to stwierdzamy, że w istocie można układowi przypisać transmitancję

$$G(s, z) = \frac{X(s, z)}{X(s, 0)} = e^{-T_0 s}; \quad T_0 = \frac{z}{v}$$

Odpowiada to dokładnie występowaniu opóźnienia czasowego

$$x(t, z) = x(t - T_0, 0)$$



·2.77· Charakterystyki układu opóźniającego: a) skokowa; b) częstotliwościowa

czyli przebieg $x(t)$ w punkcie z ma identyczny kształt jak $x(t)$ w punkcie $z = 0$, tylko opóźniony o czas $T_0 = \frac{z}{v}$. Na rysunku 2.77 przedstawiono charakterystykę skokową oraz charakterystyki częstotliwościowe członu opóźniającego. Człon taki jest często traktowany jako jeszcze jeden człon elementarny, uzupełniający zestaw przedstawiony w tabl. 2.52. Typowe jest bowiem występowanie zjawiska opóźnienia czasowego w połączeniu z elementarnymi zjawiskami „skupionymi”, np. z inercją lub całkowaniem.

2.78. Czy człon opóźnienia czasowego można w przybliżeniu przedstawić za pomocą modelu o parametrach skupionych?

Tak, ponieważ transmitancję $e^{-T_0 s}$ można przedstawić jako granicę

$$e^{-T_0 s} = \lim_{n \rightarrow \infty} \left(\frac{1}{\frac{T_0}{n} s + 1} \right)^n$$

Opóźnienie czasowe jest więc wynikiem szeregowego połączenia nieskończenie wielu członów inercyjnych o stałych czasowych dążących do zera. Widać wyraźnie, że układ ma nieskończenie wiele zmiennych stanu (wielkości charakteryzujących poszczególne inercyjne ogniwa tego nieskończonego łańcucha). Można więc sądzić, że przybliżeniem członu opóźniającego jest układ o transmitancji $\frac{1}{(T_n s + 1)^n}$ przy n dostatecznie dużym i T_n małym. Rzeczywiście, przybliżenie takie jest dość dobre, zwłaszcza jeśli w układzie występują jeszcze inne zjawiska dynamiczne, a więc np.

$$\frac{e^{-T_0 s}}{Ts + 1} \approx \frac{1}{(Ts + 1)(T_n s + 1)^n}; \quad n \gg 1, \quad T_n \ll T$$

Często stosuje się przybliżenie członu opóźniającego uzyskane dzięki rozwinięciu

$$e^{-T_0 s} = \frac{e^{-\frac{T_0}{2}s}}{e^{\frac{T_0}{2}s}} = \frac{1 - \frac{T_0}{2}s + \dots}{1 + \frac{T_0}{2}s + \dots}$$

Należy podkreślić, że transmitancja typu $e^{-T_0 s}$, a także jej aproksymacje, odpowiadają układom nieminimalnofazowym. Jest to istotna cecha układów z opóźnieniem czasowym (a także innych układów o parametrach rozłożonych). Układy te mają wyraźnie zwiększone przesunięcie fazowe w porównaniu z układami o parametrach skupionych, mających podobne charakterystyki amplitudowe (widać to wyraźnie na rys. 2.77b — charakterystyka amplitudowa jest tam stała, jak w przypadku członu proporcjonalnego).

2.79. Jak można inaczej opisać zjawisko opóźnienia czasowego?

Do opisu opóźnienia czasowego można także zastosować opis za pomocą tzw. równania różniczkowo-różnicowego lub równania z odchylnym argumentem. Jest to naturalny sposób opisu w sytuacji, gdy w układzie występuje wyraźne zjawisko **transportu**, a więc gdy sterowanie u , zanim dotrze do właściwego układu dynamicznego, podlega przesunięciu w czasie — np. gdy paliwo lub reagent dostarcza się do pieca czy reaktora chemicznego za pomocą transportera. Jeśli takie przesunięcie jest związane z czasem T_0 , a część istotnie dynamiczna układu ma np. charakter członu inercyjnego, to równanie układu można przedstawić w postaci

$$T \frac{dx}{dt} + x = u(t - T_0)$$

przy czym w równaniu tym nie wystarcza podać warunku początkowego $x(0)$, lecz trzeba podać przebieg $x(t)$, $0 \leq t < T_0$, tzn. przebieg zmiennej stanu na odcinku czasu o długości T_0 — jest to tzw. stan zupełny tego układu. Równaniu układu odpowiada transmitancja

$$\frac{X(s)}{U(s)} = e^{-T_0 s} \frac{1}{Ts + 1}$$

W ogólniejszym przypadku, gdy lewa strona równania ma bardziej rozbudowaną postać (zawiera pochodne wyższych rzędów), transmitancję układu można zapisać jako

$$\frac{X(s)}{U(s)} = e^{-T_0 s} G(s)$$

przy czym $G(s)$ odpowiada części dynamicznej układu. Jest to postać transmitancji często spotykana w obiektach przemysłowych.

·2.80· Jak opisuje się zjawisko dyfuzji?

Podstawowe równanie dyfuzji (a także przewodzenia ciepła) ma postać

$$\frac{\partial x(t, z)}{\partial t} = a \frac{\partial^2 x(t, z)}{\partial z^2}; \quad a > 0$$

Zmienna x odpowiada stężeniu dyfundującej substancji lub temperaturze. Rozwiązując to równanie w podobny sposób jak równanie przepływu w pyt. 2.77 otrzymujemy

$$a \frac{d^2 X(s, z)}{dz^2} - sX(s, z) = 0$$

Przy założeniu, że $x(t, z)$ jest ograniczone przy dowolnym z (lub co równoważne, że wymiar geometryczny z układu jest nieograniczony), rozwiązaniem równania jest

$$X(s, z) = e^{-\sqrt{\frac{s}{a}} z} X(s, 0)$$

Tak więc transmitancja układu o wejściu i wyjściu skupionym ma postać

$$G(s, z) = e^{-\sqrt{\frac{z}{a}}}$$

Otrzymany wynik różni się jakościowo od rozwiązania w przypadku przepływu, w szczególności charakterystyka skokowa nie wykazuje opóźnienia.

Należy jednak podkreślić, że obydwie pokazane transmitancje (dla przepływu i dla dyfuzji) są funkcjami **przestępnymi** zmiennej s . Transmitancję będącą funkcją wymierną zmiennej s (a więc stanowiącą iloraz wielomianów) można otrzymać wyłącznie w przypadku układu o parametrach skupionych.

·2.81· Jak opisuje się zjawiska falowe?

Równanie falowe dla najprostszego przypadku ma postać

$$\frac{\partial^2 x(t, z)}{\partial t^2} = v^2 \frac{\partial^2 x(t, z)}{\partial z^2}$$

Parametr v odpowiada prędkości propagacji fal; w przypadku kontinuum LC (linii długiej) $x(t, z)$ oznacza napięcie w chwili t w punkcie z linii, ponadto $v^2 = \frac{1}{lc}$ (l i c są parametrami elektrycznymi linii na jednostkę długości). W rozwiązaniu, które można otrzymać metodą podobną do poprzedniej, pojawiają się charakterystyczne dwa składniki odpowiadające falam „bieżącym” i „odbitym”

$$X(s, z) = [Ae^{-\frac{z}{v}} + Be^{\frac{z}{v}}] X(s, 0)$$

Odróżnia to zasadniczo rozwiązanie równania falowego od równania przepływu, choć przy braku fali odbitej ($B = 0$, $A = 1$), co ma miejsce np. przy nieograniczonej długości linii, również w rozważanym przypadku otrzymuje się opóźnienie czasowe

$$G(s, z) = \frac{X(s, z)}{X(s, 0)} = e^{-T_0 s}; \quad T_0 = \frac{z}{v}$$

Zjawiska falowe odróżniają się od zjawisk przepływu również tym, że prędkość propagacji v zależy od właściwości ośrodka, a nie jest narzucona prędkością nośnika.

·2.82· Jaka jest ogólna postać opisu układów dynamicznych nieliniowych?

Za najbardziej ogólną postać opisu układów dynamicznych nieliniowych można uznać równania różniczkowe (równania stanu). W przypadku układu o parametrach skupionych można określić n -wymiarowy wektor stanu x , r -wymiarowy wektor wejść u i m -wymiarowy wektor wyjść y , podobnie jak w przypadku liniowym. Równanie stanu można zapisać jako

$$\dot{x} = f(x, u)$$

zaś równanie wyjścia jako

$$y = g(x, u)$$

Funkcje wektorowe f i g należy rozumieć jako wektory odpowiednio n - i m -wymiarowe, których składowe są funkcjami argumentów wektorowych x i u czyli wszystkich składowych wektorów x i u . Funkcje f i g są jednak **nieliniowe**, co powoduje, że nie można stosować wygodnych wyrażeń macierzowych.

·2.83· Czy istnieje ogólna analityczna metoda rozwiązywania równań układu nieliniowego?

Nie — i jest to podstawowa trudność analizy układów nieliniowych. Nie można również stosować aparatu pojęciowego związanego z przekształceniem Laplace'a, a więc transmitancji, charakterystyk czasowych i częstotliwościowych. Nie istnieją wartości własne. Jak wiadomo z teorii równań różniczkowych, istnieją metody analityczne rozwiązywania niektórych typów równań nieliniowych. Rzadko jednak w praktyce spotyka się nieliniowości prowadzące do klasycznych typów równań nieliniowych.

·2.84· Co to są układy z nieliniowością statyki?

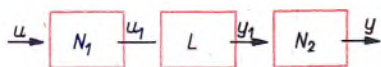
Są to układy dynamiczne nieliniowe, w których można wyodrębnić część statyczną układu opisywaną modelem nieliniowym, natomiast część dynamiczną można opisać za pomocą modelu liniowego. Z powyższego określenia nie wynika, że układ rzeczywiście składa się z dwóch części — statycznej i dynamicznej. Statykę układu można rozumieć jako zależność wiążącą wejście i wyjście układu w stanie ustalonym, tzn. przy x , y i u w przybliżeniu stałych. W przypadku skalarnym, jeśli rozważyć równanie różniczkowe wejście-wyjście o postaci

$$\begin{aligned} \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \dots + a_1 \frac{dy}{dt} + a_0 y &= \\ &= b_m \frac{d^m u}{dt^m} + \dots + b_0 u \end{aligned}$$

w którym współczynniki $a_0, \dots, a_{n-1}$ i $b_0, \dots, b_m$ zależą od współrzędnych układu (powodując nieliniowość równania) statyka jest więc określona tylko przez zależność

$$a_0 y = b_0 u$$

Jeśli tylko ta zależność jest nieliniowa, tzn. wyraża się jako $y = f_N(u)$, przy czym f_N jest pewną funkcją nieliniową, natomiast współczynniki $a_1, \dots, a_{n-1}$, $b_1, \dots, b_m$ są stałe, to mamy właśnie do czynienia z nieliniowością statyki. Ponieważ przypadek ten jest niewątpliwie najprostszym możliwym wariantem układu dynamicznego nieliniowego, więc często staramy się sprowadzić do takiej postaci (z pewnym przybliżeniem) modele układów istotnie nieliniowych. Układ



·2.84· Struktura modelu najprostszego układu z nieliniowością statyki (N_1 i N_2 — człony nieliniowe, L — człon liniowy)

z nieliniowością statyki można przedstawić tak jak na rys. 2.84. Człony nieliniowe N_1 i N_2 można opisać za pomocą zależności analitycznej lub charakterystyki statycznej, natomiast człon liniowy L scharakteryzuje transmitancja. Kolejność łączenia jest oczywiście istotna, tak samo jak i współrzędne punktu pracy na charakterystykach członów nieliniowych.

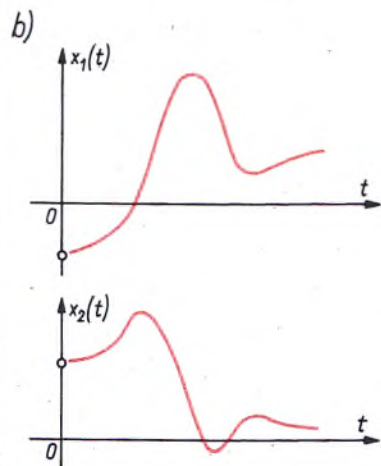
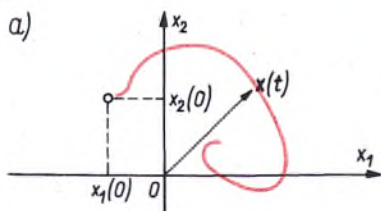
·2.85· Czy układy dynamiczne nieliniowe można linearyzować?

Tak, badając układ nieliniowy przy założeniu małych odchyśleń od pewnego punktu pracy układu (np. jego stanu równowagi), można rozwinąć funkcje nieliniowe w szereg Taylora, pominąć wyrazy wyższych rzędów i otrzymać w ten sposób równania przybliżone liniowe. Należy jednak przypomnieć zastrzeżenia co do możliwości linearyzacji, wymienione w pyt. 2.26, z których wynika, że nie każdy układ nieliniowy można poddać linearyzacji. Może się także okazać, że nie istnieje stan równowagi, wokół którego można by dokonać rozsądnej linearyzacji. Oczywiście szczególnie podatne dla idei linearyzacji są układy z nieliniowością części statycznej.

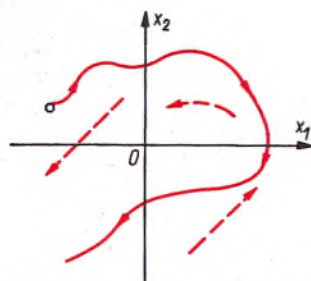
·2.86· W jaki sposób można przedstawić rozwiązania równań stanu układu nieliniowego?

Rozwiązanie równań stanu układu — jeśli oczywiście to rozwiązanie można uzyskać — ma postać tzw. **trajektorii** stanu układu. Trajektorię można sobie wyobrazić jako pewną linię skierowaną w n -wymiarowej przestrzeni stanów, zaczynającą się od stanu początkowego $x(0)$, a w kolejnych (bieżących) chwilach czasu t określonych aktualnymi współrzędnymi wektora stanu $x(t)$. W geometrycznej interpretacji wektora stanu trajektoria jest linią zakresloną przez koniec wektora stanu w trakcie „ruchu” układu, przy czym czas t jest parametrem trajektorii.

Na rysunku 2.86a pokazano trajektorię w przypadku dwuwymiarowym, kiedy przestrzeń stanów sprowadza się do płaszczyzny o współrzędnych x_1 i x_2 , a wektor stanu można przedstawić jako zwykły wektor na płaszczyźnie. Przebiegi czasowe $x_1(t)$ i $x_2(t)$ (rys. 2.86b) odpowiadają trajektorii z rys. (a). Chociaż zazwyczaj za rozwiązania równań różniczkowych uważa się zależności określające przebiegi czasowe, to jednak — zwłaszcza dla układów nieliniowych — pożądaną i poglądową postacią rozwiązania jest zależność określająca trajektorię. Oczywiście przebiegi czasowe wszystkich zmiennych stanu $x_1(t), \dots, x_n(t)$ określają całkowicie trajektorię, jednak bardziej interesujące może się okazać wyeliminowanie zmiennej t i opisanie trajektorii za pomocą zależności wiążącej wartości wszystkich zmiennych stanu w dowolnych chwilach czasu. Najlepsze rezultaty daje tzw. **metoda płaszczyzny fazowej**.



·2.86· Rozwiązania równań stanu w przypadku układu drugiego rzędu: a) trajektoria stanu; b) przebiegi czasowe



·2.87· Typowy kształt trajektorii fazowej; linie przerywane — trajektorie niedopuszczalne

·2.87· Na czym polega metoda płaszczyzny fazowej?

Metoda płaszczyzny fazowej jest szczególnym przypadkiem zastosowania opisu w przestrzeni stanów. Metodę płaszczyzny fazowej stosuje się do układów drugiego rzędu, których zmienne stanu x_1 i x_2 określają współrzędne prostokątne punktu na płaszczyźnie; ponadto x_1 i x_2 są współzrędnymi fazowymi, tzn. $\dot{x}_1 = x_2$. Wobec tego metodę płaszczyzny fazowej stosuje się do układów opisanych równaniami stanu

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = f(x_1, x_2)$$

Metoda ta wykorzystuje szczególną łatwość określenia i interpretacji **trajektorii fazowej**, czyli trajektorii stanu w tych szczególnych współrzędnych. Trajektorja jest tu bowiem zwykłą linią (krzywą) na płaszczyźnie i jej cechy geometryczne można bardzo łatwo określić. W szczególności, z określenia współrzędnych fazowych wynika, że można im nadać sens „położenia” (x_1) i „prędkości” ($x_2 = \dot{x}_1$). Zatem wartość zmiennej x_1 przy $x_2 > 0$ powinna wzrastać, przy $x_2 < 0$ — maleć, a przy $x_2 = 0$ — osiągać lokalne ekstremum. Tak więc trajektorie fazowe przebiegają w górnej półpłaszczyźnie w prawo, w dolnej półpłaszczyźnie w lewo, oś x_1 mogą przecinać, ale ze styczną prostopadłą do tej osi (rys. 2.87).

Pomimo swej ograniczoności do układów drugiego rzędu metoda płaszczyzny fazowej jest powszechnie stosowana. Wynika to z dwóch przyczyn: po pierwsze — układy drugiego rzędu (i pierwszego) wyczerpują w zasadzie podstawowe typy dynamiki układów — w szczególności obejmują człon oscylacyjny wykazujący największą różnorodność zjawisk. Po drugie — jeśli nawet w rzeczywistości układ nieliniowy jest wyższego rzędu, to w większości przypadków, zwłaszcza dla dużych zmian sygnałów, można go aproksymować układem rzędu drugiego; w każdym razie jest to rozsądne „pierwsze przybliżenie” układu nieliniowego wyższego rzędu niż dwa.

·2.88· Co to jest portret fazowy układu?

Jest to rodzina trajektorii fazowych przy różnych warunkach początkowych. Pojedyncza trajektorja zawiera niewiele informacji o układzie. Metoda płaszczyzny fazowej ujawnia swoje największe zalety wtedy, gdy można przedstawić wiele trajektorii fazowych przy różnych warunkach początkowych. Jeśli tak utworzona rodzina trajektorii pokryje dość gęsto całą płaszczyznę, to otrzymujemy wtedy obraz wszystkich potencjalnych rozwiązań przy dowolnych warunkach początkowych. Oczywiście każdy punkt na płaszczyźnie fazowej może być warunkiem początkowym, strzałki na trajektoriach (zgodnie z zasadami wymienionymi w poprzednim pytaniu) wskazują kierunek upływu czasu. W tej sytuacji

portret fazowy faktycznie dostarcza pełnej informacji o właściwościach układu.

Zwróćmy uwagę, że z jednoznaczności rozwiązań równania różniczkowego wynika jednoznaczność płaszczyzny fazowej, tzn. przez dany punkt płaszczyzny fazowej może przechodzić tylko jedna trajektoria — a więc trajektorie nie mogą się przecinać. Wyjątkiem od tej zasady są tzw. punkty osobliwe, w których nie można określić równania trajektorii.

•2.89• Jak określa się równanie trajektorii i kiedy pojawiają się punkty osobliwe?

Mając dane równania stanu

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = f(x_1, x_2)$$

można podzielić drugie z równań przez pierwsze, otrzymując

$$\frac{dx_2}{dx_1} = \frac{f(x_1, x_2)}{x_2}$$

Jest to równanie różniczkowe trajektorii fazowej, którego rozwiązanie stanowi rodzina krzywych $x_2 = x_2(x_1)$ odpowiadających różnym warunkom początkowym (x_1^0, x_2^0) . Równanie trajektorii nie jest określone, jeśli zachodzi równość

$$\dot{x}_1 = \dot{x}_2 = 0$$

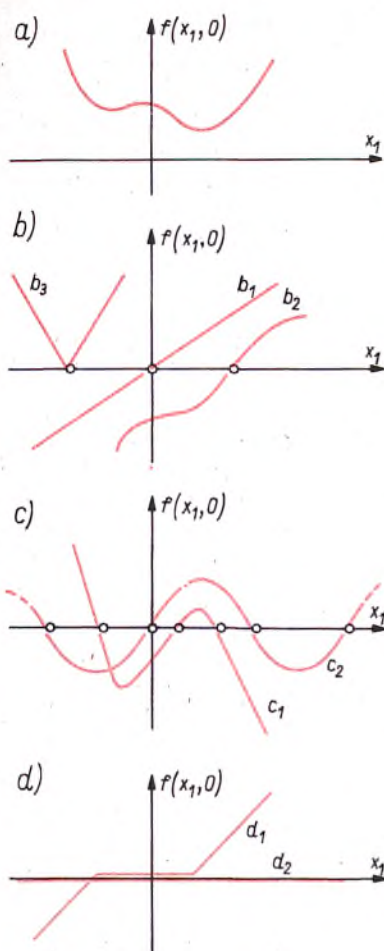
Warunek ten określa **punkty osobliwe**, w których pochodna dx_2/dx_1 jest nieoznaczona, zatem przez punkt osobliwy może przechodzić więcej niż jedna trajektoria.

Równanie różniczkowe trajektorii jest z zasady łatwiejsze do rozwiązania niż równania stanu. Co więcej, w wyniku rozwiązania równania trajektorii otrzymuje się bezpośrednie wyrażenie trajektorii jako zależności $x_2(x_1)$ — natomiast po rozwiązaniu równań stanu otrzymuje się $x_1(t)$ i $x_2(t)$, skąd samą trajektorię można otrzymać dopiero po wyeliminowaniu zmiennej t . Jednak w wielu przypadkach równanie trajektorii może być także dość skomplikowane. Korzysta się wtedy z przybliżonych metod rozwiązania, z których najbardziej znaną jest **metoda izoklin**. Izokliną nazywa się linię $\frac{dx_2}{dx_1} = A = \text{const}$, tzn. linię, na której nachylenie trajektorii jest stałe. Równanie izokliny jest równaniem algebraicznym (a nie różniczkowym) i ma postać

$$\frac{f(x_1, x_2)}{x_2} = A$$

Zmieniając stałą A otrzymuje się różne izokliny — z określonym nachyleniem A trajektorii.

•2.90• Ile punktów osobliwych może występować na płaszczyźnie fazowej?



·2.90· Położenie punktów osobiwych na płaszczyźnie fazowej; a) brak punktów osobiwych; b) jeden punkt osobiwy; c) skończona lub przeliczalna liczba punktów osobiwych; d) kontinuum punktów osobiwych

Ponieważ warunkiem występowania punktów osobiwych jest równość $\dot{x}_1 = \dot{x}_2 = 0$, więc można stąd wyprowadzić wnioski bardziej praktyczne:

- punkty osobiwe mogą leżeć tylko na osi x_1 (ponieważ $\dot{x}_1 = x_2 = 0$);
- położenie punktów osobiwych na osi x_1 jest określone równaniem wynikającym z równości $\dot{x}_2 = 0$

$$f(x_1, x_2 = 0) = 0$$

Punkty osobiwe są więc określone przez miejsca zerowe prawej strony drugiego z równań stanu względem zmiennej x_1 przy $x_2 = 0$. Mogą zaistnieć następujące przypadki zilustrowane na rys. 2.90.

a) Nie ma punktów osobiwych, jeśli $f(x_1, 0)$ jest stale jednego znaku; jest to możliwe tylko w przypadku nieliniowym.
b) Jeden punkt osobiwy; jest to typowe dla układów liniowych, w których obowiązuje równanie $f(x_1, x_2) = -a_0 x_1 - a_1 x_2$, tzn. punkt osobiwy leży zawsze w początku układu współrzędnych $(0, 0)$; w przypadkach nieliniowych może także występować pojedynczy punkt osobiwy — niekoniecznie $(0, 0)$; szczególnym (zdegenerowanym) przypadkiem jest przypadek, gdy charakterystyka statyczna $f(x_1, 0)$ styka się tylko z osią x_1 , lecz jej nie przecina (krzywa b_3).

c) Skończona (większa niż jeden) lub przeliczalna liczba punktów osobiwych; jest to możliwe tylko w układach nieliniowych (liczba skończona punktów osobiwych wystąpi np. przy $f(x_1, 0)$ o postaci wielomianu, liczba nieskończona lecz przeliczalna — np. przy $f(x_1, 0) = \sin x_1$).

d) Zbiór nieprzeliczalny (kontinuum) punktów osobiwych; jest to możliwe w przypadku nieliniowym, jeśli istnieje cały przedział (lub przedziały) wartości x_1 , w którym obowiązuje równość $f(x_1, 0) = 0$ (np. charakterystyka typu strefy nieczułości; zbiór nieprzeliczalny może wystąpić także w przypadku liniowym, gdy drugie z równań stanu nie zależy od x_1 , co wiąże się z zerowością przynajmniej jednej wartości własnej układu i co powoduje, że cała oś x_1 stanowi zbiór punktów osobiwych.

·2.91· Jakie rozróżnia się typy punktów osobiwych?

Zauważmy najpierw, że punkty osobiwe są to jedyne stany równowagi układu, bowiem warunek $\dot{x}_1 = \dot{x}_2 = 0$ można interpretować jako warunek spoczynku (bezruchu) układu. Równowaga ta może być trwała lub nietrwała, wobec tego podstawowy podział polega na rozróżnieniu stabilnych i niestabilnych punktów równowagi. Nieścista, lecz pogładowa ocena stabilności punktu równowagi sprowadza się do stwierdzenia, czy wszystkie trajektorie mają kierunek do danego punktu (wtedy jest on stabilny), czy też — przynajmniej niektóre trajektorie — mają kierunek od punktu (wtedy jest on niestabilny). Dokładniejszej klasyfikacji punktów równowagi dokonuje się na podstawie przybliżenia liniowego drugiego z równań stanu w sąsiedztwie punktu równowagi

(oczywiście tylko dla przypadków, w których takie przybliżenie ma sens). Rozważmy układ liniowy o równaniach stanu:

$$\dot{x}_1 = x_2$$

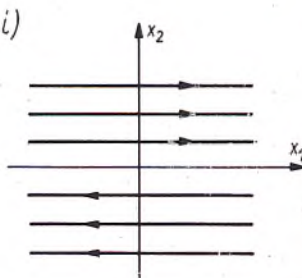
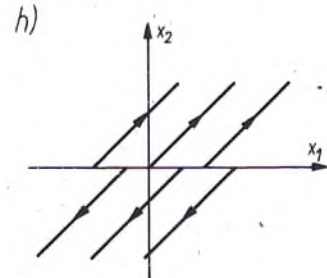
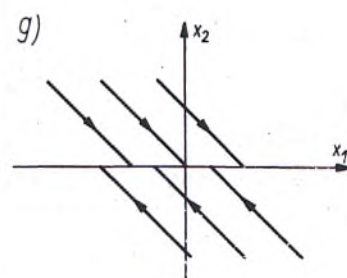
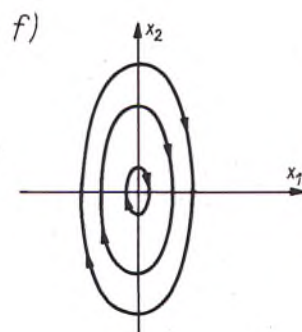
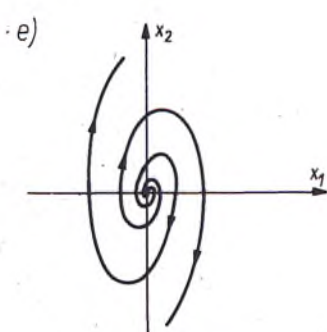
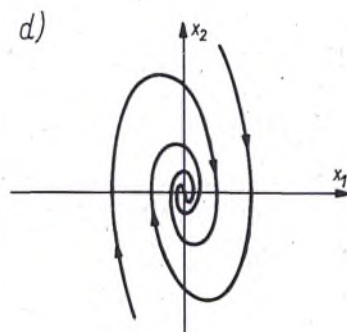
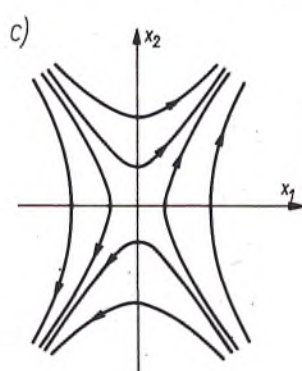
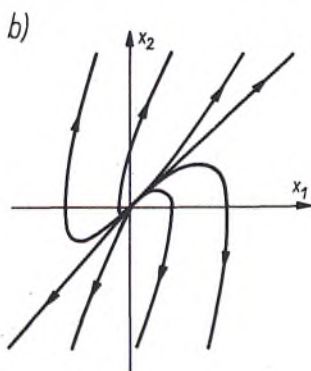
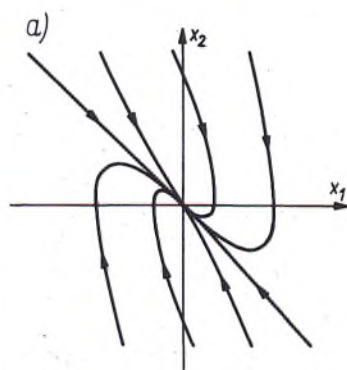
$$\dot{x}_2 = -a_0 x_1 - a_1 x_2$$

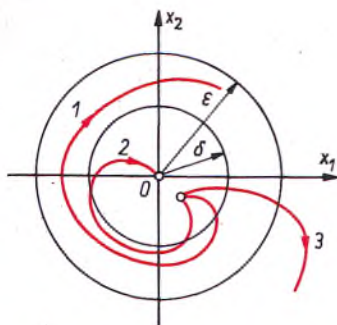
Wobec tego ma on dwie wartości własne spełniające równanie charakterystyczne

$$\lambda^2 + a_1 \lambda + a_0 = 0$$

Szybkie oszacowanie wartości własnych λ_1 i λ_2 ułatwia spostrzeżenie, że $\lambda_1 \lambda_2 = a_0$, zaś $\lambda_1 + \lambda_2 = -a_1$. Typ punktu osobiwego (punktu równowagi) jest określony przez charakterystyczne kształty trajektorii, związane z różnymi kom-

-2.91• Typy punktów osobiwych układu liniowego i odpowiednie portrety fazowe: a) węzeł stabilny; b) węzeł niestabilny; c) siodło; d) ognisko stabilne; e) ognisko niestabilne; f) środek; g) $s_1 = 0, s_2 < 0$ — punkty osi x_1 stabilne; h) $s_1 = 0, s_2 > 0$ — punkty osi x_1 niestabilne; i) $s_1 = s_2 = 0$ — punkty osi x_1 niestabilne





•2.92• Ilustracja definicji stabilności punktu równowagi w sensie Lapunowa:

1 — stabilność; 2 — stabilność asymptotyczna; 3 — niestabilność

binacjami wartości własnych. Na rysunku 2.91 pokazano wszystkie możliwe przypadki:

I. $a_1^2 \geq 4a_0$ (wartości własne rzeczywiste)

a) $\lambda_1 < 0, \lambda_2 < 0$, punkt osobliwy nazywa się **węzłem stabilnym**;

b) $\lambda_1 > 0, \lambda_2 > 0$ — **węzłem niestabilnym**;

c) $\lambda_1 \lambda_2 < 0$ — **siodło**.

II. $a_1^2 < 4a_0$ (wartości własne zespolone sprzężone)

d) $\text{Re } \lambda_1 = \text{Re } \lambda_2 < 0$ — **ognisko stabilne**;

e) $\text{Re } \lambda_1 = \text{Re } \lambda_2 > 0$ — **ognisko niestabilne**

f) $\text{Re } \lambda_1 = \text{Re } \lambda_2 = 0$ — tzn. wartości własne urojone — **środek**.

W przypadkach jednej lub obu wartości własnych zerowych otrzymuje się kontinuum punktów równowagi na osi x_1 i portrety fazowe pokazane na rys. 2.91g, h, i.

•2.92• Jak określa się w sposób ścisły stabilność punktu równowagi?

Najbardziej powszechna jest definicja stabilności punktu równowagi w sensie Lapunowa. Zgodnie z tą definicją, jeśli dla każdego (dowolnie małego) obszaru o promieniu ε , otaczającego badany punkt równowagi, można znaleźć taki obszar o promieniu $\delta(\varepsilon)$, że trajektorie, zaczynające się z punktów określonych warunkami początkowymi i leżących wewnątrz obszaru o promieniu δ , nie wychodzą poza obszar o promieniu ε , to punkt równowagi jest stabilny. Jeśli ponadto, przy czasie dążącym do nieskończoności, trajektorie dążą do samego punktu równowagi, to punkt ten jest **stabilny asymptotycznie**. Na rysunku 2.92 przedstawiono interpretację definicji Lapunowa na płaszczyźnie fazowej.

Podana definicja umożliwia uzasadnienie typu stabilności wszystkich typów punktów równowagi na płaszczyźnie fazowej: w przypadku $\text{Re } \lambda_{1,2} < 0$ występują zawsze punkty równowagi stabilnej; przy $\text{Re } \lambda_i > 0$ mamy do czynienia z niestabilnością (siodło też jest punktem niestabilnym); środek jest punktem stabilnym, lecz nie stabilnym asymptotycznie, podobnie punkty osi x_1 przy $\lambda_1 = 0, \lambda_2 < 0$; wreszcie przy $\lambda_1 = \lambda_2 = 0$ punkty osi x_1 są niestabilne. Oczywiście pojęcia punktu równowagi (punktu osobliwego) i jego stabilności odnoszą się do układów dowolnego rzędu (skończonego). Punkty równowagi określa się ogólnie z równania

$$\dot{x} = 0$$

•2.93• Czy określenie stabilności w sensie Lapunowa dotyczy tylko otoczenia punktu równowagi?

Tak, w zasadzie definicja ta określa stabilność **lokalną**. Dla rozszerzenia sensu definicji wprowadza się, w przypadku stabilności asymptotycznej, pojęcie obszaru stabilności jako obszaru warunków początkowych, z których trajektorie dążą

asymptotycznie do punktu równowagi. Jeśli obszar stabilności jest nieograniczony, to mówi się, że punkt równowagi jest stabilny **globalnie**. Jest to najsilniejsze określenie stabilności i odpowiada bez zastrzeżeń sensowi intuicyjnemu pojęcia stabilności jako właściwości „powrotu do punktu równowagi” przy dowolnych zaburzeniach.

Określenie stabilności punktu równowagi zazwyczaj rozciąga się na układ, a więc o układzie, którego opis w przestrzeni stanów (w szczególności na płaszczyźnie fazowej) wykazuje istnienie stabilnego punktu (lub punktów) równowagi, mówi się, że jest stabilny.

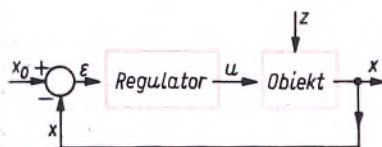
3.1. Jak można przedstawić strukturę układu regulacji?

Jak już powiedziano na wstępie (rozdz. 1), regulacja jest to sterowanie w układzie zamkniętym, a więc w układzie ze sprzężeniem zwrotnym. W związku z tym każdy układ regulacji automatycznej można przedstawić za pomocą schematu blokowego pokazanego na rys. 3.1A.

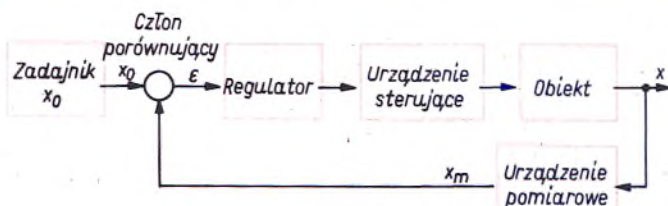
Podstawowymi członami układu regulacji są oczywiście: **obiekt regulacji** oraz **regulator**. Podstawowymi sygnałami w układzie regulacji są natomiast: **sygnał regulowany** x , **sygnał wartości zadanej** x_0 , **sygnał uchybu** ϵ oraz **sygnał sterujący (nastawiający)** u , a także **sygnał zakłócający** z . Schemat pokazany na rys. 3.1A jest wyidealizowany, ponieważ w praktyce składa się on jeszcze z innych członów (i sygnałów), wynikających z oprzyrządowania układu regulacji. Na rysunku 3.1B pokazano trochę bardziej rozwinięty schemat uwzględniający oprzyrządowanie. Poza właściwym regulatorem można odróżnić oprzyrządowanie pomiarowe, oprzyrządowanie sterownicze i zadajnik wartości zadanej. Nie wymieniono członów sygnalizacyjnych, zabezpieczających, umożliwiających manipulacje ręczne itd. urządzenia te są omówione dokładniej w rozdz. 4.

3.2. Jaka jest rola poszczególnych sygnałów w układzie regulacji?

Członem pierwotnym struktury układu regulacji jest obiekt regulacji. Obiekt jest **dany** w tym sensie, że jest to przedmiot, którego struktury wewnętrznej zmienić nie można, natomiast można wytypować pewną wielkość charakteryzującą



3.1A. Ogólny schemat blokowy układu regulacji



3.1B. Rozwinięty schemat blokowy układu regulacji

działanie obiektu, którą to wielkość zamierzamy regulować. Mówiąc o regulowaniu, mamy na myśli nadawanie wielkości regulowanej pożądanych wartości lub przebiegów — określonych przez sygnał wartości zadanej. Sens stosowania regulacji wynika z obecności zakłóceń — gdyby ich nie było, nie byłoby powodu budować układu regulacji. Sygnał sterujący określa, czym można oddziaływać na obiekt dla osiągnięcia zamierzonego celu. Wreszcie uchyb regulacji stanowi elementarną miarę spełnienia postawionego zadania. W idealnym układzie regulacji **uchyb powinien być stale równy zeru**; właściwość ta ma zupełnie zasadnicze znaczenie, można ją uważać za cechę definicyjną układu regulacji. Istotne jest przy tym, aby dążenie do zlikwidowania uchybu miało miejsce nawet przy niewielkiej informacji początkowej o obiekcie, zakłóceniach i ewentualnych zmianach właściwości obiektu.

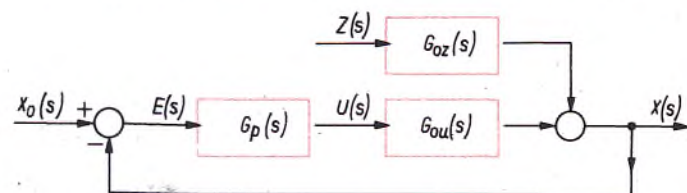
•3.3• Jakież zależności występują między sygnałami w układzie regulacji?

Przyjmijmy na razie, że rozpatrujemy układy liniowe, wobec czego układ regulacji opiszemy za pomocą transmitancji. Schemat blokowy układu regulacji przedstawiono na rys. 3.3. Transmitancja obiektu regulacji względem sygnału sterującego u jest oznaczona przez $G_{ou}(s)$ oraz przez $G_{oz}(s)$ względem zakłócenia z ; $G_r(s)$ oznacza transmitancję regulatora. W szczególnym przypadku mogą zachodzić równości $G_{oz}(s) = 1$ (zakłócenie na wyjściu obiektu) lub $G_{oz}(s) = G_{ou}(s)$ (zakłócenie na wejściu obiektu). Przez $G_0(s)$ oznaczmy **transmitancję układu otwartego** regulacji, łączącą uchyb z wielkością regulowaną (transformaty $E(s)$ z $X(s)$): $G_0(s) = G_r(s)G_{ou}(s)$. Łatwo sprawdzić słuszność następujących zależności:

$$\begin{aligned} X(s) &= \frac{G_0(s)}{1+G_0(s)} X_0(s) + \frac{G_{oz}(s)}{1+G_0(s)} Z(s) = \\ &= G(s)X_0(s) + G_z(s)Z(s) \end{aligned}$$

$$\begin{aligned} E(s) &= \frac{1}{1+G_0(s)} X_0(s) - \frac{G_{oz}(s)}{1+G_0(s)} Z(s) = \\ &= G_e(s)X_0(s) - G_z(s)Z(s) \end{aligned}$$

Wyrażenie $G(s) = G_0(s)/(1+G_0(s))$ nazywa się **transmitancją układu zamkniętego** regulacji, ponieważ określa ono związek pomiędzy x_0 i x . Największe znaczenie ma jednak



•3.3• Schemat blokowy liniowego modelu układu regulacji

wielkość $G_e(s) = 1/(1+G_0(s))$, określająca **transmitancję uchybową**. Wyrażenie to ma ponadto sens czynnika określającego wpływ wprowadzenia sprzężenia zwrotnego (tj. regulacji). Przez ten czynnik są bowiem w podanych zależnościach pomnożone transmitancje układu otwartego, a więc $G(s) = G_0(s) \cdot G_e(s)$ oraz $G_z(s) = G_{oz}(s)G_e(s)$ — ta ostatnia transmitancja odnosi się do wpływu zakłóceń w układzie zamkniętym i jest równa G_e przy $G_{oz} = 1$ (przy zakłóceniach na wyjściu obiektu), natomiast przy zakłóceniach na wejściu obiektu ($G_{oz} = G_{ou}$) otrzymuje się $G_z(s) = G(s)/G_r(s)$.

·3.4· Jak można sklasyfikować zadania stawiane układowi regulacji?

Przedstawione w poprzednim pytaniu zależności sugerują możliwe podziały zadań układów regulacji w zależności od relacji wiążących sygnały $x(t)$ i $\varepsilon(t)$ z jednej strony — z sygnałami $x_0(t)$ i $z(t)$ z drugiej strony. Cel układu regulacji określa w zasadzie sygnał zadany x_0 — stąd podział rodzajów regulacji ze względu na postać sygnału x_0 jest następujący:

— Regulacja stałowartościowa przy $x_0 = \text{const}$.
 — Regulacja programowa przy sygnale $x_0(t)$ będącym funkcją czasu daną z góry jako pewien „program” zmian wielkości regulowanej.

— Regulacja nadążna przy $x_0(t)$ w postaci przebiegu o nieznanym z góry kształcie (np. przebiegu przypadkowego). Podział ten jest jednak niedoskonały, ponieważ ignoruje rolę zakłócenia i nie uwypatnia najważniejszych problemów projektowych. Z tego względu zadania układów regulacji można podzielić następująco:

— Zadania nadążania, jeśli układ jest w stanie spełnić przybliżoną równość $\varepsilon \approx 0$ w każdej chwili, przy określonych klasach sygnałów x_0 i z ; powstaje wówczas problem określenia i osiągnięcia zadowalająco małych uchybów.

— Zadania przestawiania, jeśli sygnały x_0 lub z wprowadzają tak silne oddziaływania (w szczególności podlegają dużym zmianom skokowym), że można z góry powiedzieć, że uchyb będzie się znacznie różnił od zera w chwili pojawienia się tych oddziaływań; powstaje wówczas problem czasu, po którym można sprowadzić uchyb w pobliże zera.

— Zadania kompensacji zakłóceń, jeśli podstawową trudnością dla układu regulacji jest wyeliminowanie wpływu zakłóceń w jak najszerszym pasmie częstotliwości.

·3.5· Jak można sklasyfikować układy regulacji?

Układy regulacji można klasyfikować uwzględniając różne kryteria. Za najmniej istotne uznajemy w tym miejscu podziały ze względu na rodzaj wielkości regulowanej (regulatory napięcia, ciśnienia, temperatury, prędkości kątowej itd.) oraz ze względu na technikę realizacji członów regulacyjnych, wykonawczych i pomiarowych (elektroniczne, pneumatyczne, hydrauliczne). Z punktu widzenia teorii regulacji bardziej

istotny jest podział uwzględniający specyfikę projektowania i zastosowań. Pod tym względem ważne jest wyróżnienie dwu dużych klas układów regulacji, nie zawierających zresztą wszystkich możliwych przypadków. Jedną z tych klas to tzw. układy **regulacji przemysłowej** (regulacji procesów wolnozmiennych — pieców, reaktorów chemicznych itp.). Druga szeroka klasa to **serwomechanizmy** — układy regulacji nadejście położenia mechanicznego.

Pod względem sposobu realizacji regulatorów układy regulacji można podzielić na następujące grupy:

— Układy regulacji ciągłej; są to najbardziej rozpowszechnione układy regulacji, w których zastosowano elementy pracujące w sposób ciągły w czasie i mające ciągłe charakterystyki statyczne.

— Układy regulacji impulsowej; w układach tych występują (nie wyłącznie) sygnały zdyskretyzowane w czasie; obecnie ta grupa układów wykazuje szybki rozwój spowodowany zastosowaniem urządzeń cyfrowych jako regulatorów dyskretnych.

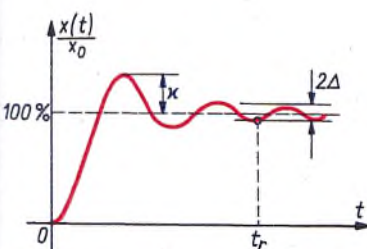
— Układy regulacji przekątnikowej; ich cechą szczególną jest zastosowanie elementów przekątnikowych o nieciągłych charakterystykach statycznych (najczęściej o dwu lub trzech wyróżnionych stanach).

Taki podział jest przydatny przede wszystkim z punktu widzenia metod analizy: układy ciągłe staramy się w miarę możliwości przedstawiać za pomocą modeli liniowych i stosować opisy transmitancyjne, w szczególności częstotliwościowe. Układy impulsowe staramy się podobnie opisywać transmitancjami dyskretnymi. Układy przekątnikowe są z zasady nieliniowe, ich analiza w małym stopniu pozwala na uproszczenia — najlepsze rezultaty otrzymuje się przez badanie trajektorii stanu, przede wszystkim metodą płaszczyzny fazowej.

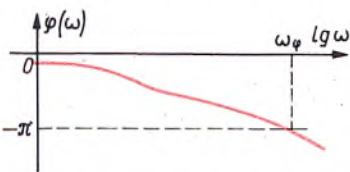
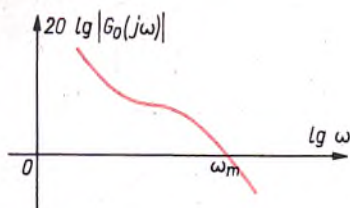
Pod względem liczby wielkości regulowanych układy regulacji można podzielić na jednowymiarowe (z jedną zmienną regulowaną) i wielowymiarowe.

•3.6. Jakie wskaźniki charakteryzują proces regulacji?

Najłatwiej można sprecyzować wymagania statyczne, czyli żądania dokładności w stanie ustalonym, ponieważ kryterium oceny jest wtedy oczywiste: **wartość uchybu ustalonego** (wyrażona ewentualnie w jednostkach względnych, w odniesieniu do wartości zadanej x_0). Ocena przebiegów przejściowych jest znacznie mniej oczywista i zależy od typu sygnału zadanego x_0 , który ma być odtwarzany przez układ regulacji. Dość powszechnie przyjęło się operowanie sygnałem skokowym jako standardowym sygnałem próbnym (x_0 lub z), przy którym bada się reakcję układu regulacji. Na tej podstawie określa się tzw. bezpośrednie wskaźniki jakości odpowiedzi skokowej. Są to przede wszystkim **czas regulacji** t_r i **przeregulowanie** $\varkappa$ — ich określenie wynika z rys. 3.6A. Czas regulacji zależy oczywiście od przyjętego marginesu Δ

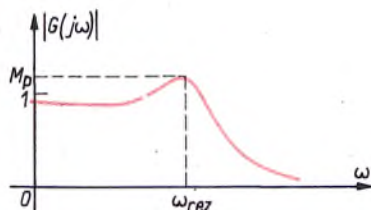


•3.6A. Określanie wskaźników bezpośrednich na podstawie odpowiedzi skokowej



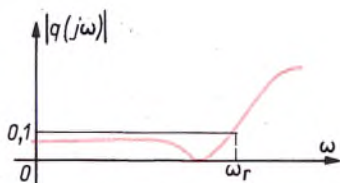
•3.6B• Określanie częstotliwości (pulsacji) granicznych na podstawie charakterystyk układu otwartego regulacji:

ω_m — pulsacja graniczna modułu, ω_φ — pulsacja graniczna fazy



•3.6C• Określanie wskaźników częstotliwościowych na podstawie charakterystyk układu zamkniętego:

ω_{rez} — pulsacja rezonansowa, M_p — moduł rezonansowy



•3.6D• Częstotliwościowy wskaźnik regulacji $q(j\omega)$ i określenie pasma roboczego z warunku np. $|q(j\omega)| \leq 0,1$

(zwykle od 1 do 5%), ponieważ w większości przypadków odpowiedź na wymuszenie skokowe osiąga stan ustalony w sposób asymptotyczny. Wskaźnikami bezpośrednimi są również wskaźniki związane z charakterystykami częstotliwościowymi, ponieważ wymuszenia sinusoidalne można również uważać za pewien typ sygnałów standardowych. Najważniejsze ze wskaźników częstotliwościowych dotyczą **częstotliwości granicznych**, określanych w układzie otwartym o transmitancji widmowej $G_0(j\omega)$ (rys. 3.6B) lub w układzie zamkniętym (rys. 3.6C), przy czym w układzie zamkniętym istotne znaczenie ma wartość **modułu rezonansowego** M_p . Częstotliwości graniczne stanowią oszacowanie **pasma roboczego regulacji** ω_r , w którym układ regulacji można uważać za skuteczny. Istnieje związek pomiędzy wskaźnikami występującymi przy odpowiedzi skokowej (t_r, κ) a wskaźnikami częstotliwościowymi (ω_r, M_p); czas regulacji stanowi w przybliżeniu odwrotność szerokości pasma regulacji, wyrażonego w Hz (tzn. $f_r = \frac{\omega_r}{2\pi}$): $t_r \approx \frac{1}{f_r}$; przeregulowanie jest tym większe, im większy jest moduł rezonansowy — dla $1 < M_p < 2$ wartość κ zależy w przybliżeniu liniowo od M_p i zmienia się od 0 do 50%. Dużą wygodę stanowi wprowadzenie tzw. **częstotliwościowego wskaźnika jakości regulacji** $q(j\omega) = \frac{1}{1+G_0(j\omega)} = G_e(j\omega)$. Idealną regulację oznaczałby warunek $|q(j\omega)| \equiv 0$, natomiast w przypadku realnym można osiągnąć przybliżoną równość $|q(j\omega)| \approx 0$ w pasmie regulacji $0 < \omega < \omega_r$ (rys. 3.6D). Inną grupę wskaźników jakości regulacji, określanych z odpowiedzi skokowej, są **wskaźniki całkowe**. Najprostsze z tych wskaźników mają postać

$$I_1 = \int_0^{\infty} \varepsilon(t) dt \quad (\text{dla } \varepsilon \text{ aperiodycznego})$$

$$I_{1m} = \int_0^{\infty} |\varepsilon(t)| dt$$

$$I_2 = \int_0^{\infty} \varepsilon^2(t) dt$$

Aby całki były zbieżne, uchyb ustalony $\varepsilon(t)$ powinien dążyć do zera dla $t \rightarrow \infty$. Wskaźniki całkowe mają tę wadę, że ich postać jest arbitralna i nie zawsze ma sens techniczny lub ekonomiczny, natomiast zaletą wskaźników jest możliwość prowadzenia sformalizowanej optymalizacji (patrz rozdz. 6). Szczególne znaczenie ma wskaźnik „kwadratowy” I_2 . Z jednej strony, ta właśnie postać umożliwia w największym stopniu uzyskanie formalnego rozwiązania zadania optymalizacji — tzn. takiego doboru struktury i parametrów układu regulacji, aby w danych warunkach wartość I_2 była minimalna (patrz rozdz. 6), a z drugiej strony można temu

wskaźnikowi przypisać interpretację dość realną: całka kwadratu pewnego sygnału ma często sens energii; całkowanie odpowiada uśrednianiu, zaś wartość średnia kwadratu w przypadku przebiegów okresowych stanowi podstawę określenia wartości skutecznej (tak jak w elektrotechnice); w przypadku sygnałów stochastycznych (losowych) średnia kwadratu sygnału może stanowić ocenę wariancji, tj. miary „rozproszenia” sygnału względem wartości przeciętnej (np. zerowej). W każdym razie należy zauważyć, że wskaźnik I_2 jest pewną miarą wypadkowego odchylenia przebiegu $\varepsilon(t)$ od wartości 0, odchylenia rozumianego w sensie energetycznym lub w sensie statystycznym. Idealny układ regulacji miałby oczywiście $I_2 = 0$, można więc sądzić, że im mniejszą wartość będzie można nadać wskaźnikowi I_2 w danym układzie regulacji, tym ten układ będzie lepszy.

·3.7· Czy jest możliwe całkowite wyeliminowanie uchybu regulacji?

Nie, ponieważ praktycznie wszystkie potencjalne obiekty regulacji (zwłaszcza jeśli włączyć do obiektu regulacji urządzenia pomiarowe i sterownicze) są układami dynamicznymi. Gdyby obiekt regulacji był układem bezinercyjnym, można by sprowadzić uchyb do wartości niemal dowolnie bliskiej zera w każdej chwili czasu. W przypadku obiektu dynamicznego sterowanie tego obiektu wymaga czasu, jeśli więc uchyb z jakichkolwiek przyczyn się pojawi, nie można go sprowadzić do zera natychmiast, co najwyżej po odpowiednio długim czasie można w pewnych warunkach osiągnąć zerowość uchybu.

·3.8· Co oznaczają pojęcia statyzmu lub astatyzmu układu regulacji?

Są to pojęcia stosowane przy ocenie układów regulacji pod względem zachowania się uchybu w stanie ustalonym. Jeśli zakłócenia są równe zera lub wchodzą na wyjście obiektu regulacji ($G_{oz} = 1$), to transformata uchybu jest określona wyrażeniem

$$E(s) = \frac{1}{1 + G_0(s)} X_0(s)$$

(zamiast $X_0(s)$ może występować $-Z(s)$). Przyjmijmy, że układ regulacji jest stabilny, a więc po dostatecznie długim czasie wartość uchybu $\varepsilon(t)$ ustala się jako ε_u . Na podstawie twierdzenia granicznego rachunku operatorowego można określić ε_u

$$\varepsilon_u = \lim_{t \rightarrow \infty} \varepsilon(t) = \lim_{s \rightarrow 0} s E(s) = \lim_{s \rightarrow 0} \frac{s}{1 + G_0(s)} X_0(s)$$

Układem statycznym regulacji nazywamy układ, w którym powyższa granica dla skokowego sygnału $x_0(t)$ (lub z) jest różna od zera i równa ε_s . Ponieważ przy $x_0(t) = A_0 \cdot 1(t)$

otrzymuje się $X_0(s) = \frac{A_0}{s}$, więc wartość ε_s , nazywaną **uchy-
bem statycznym** można obliczyć

$$\varepsilon_s = \lim_{s \rightarrow 0} \frac{A_0}{1 + G_0(s)} = A_0 \frac{1}{1 + k_0}$$

przy czym $k_0 = G_0(0)$ jest współczynnikiem wzmocnienia statycznego (wzmocnienia w stanie ustalonym) — pod warunkiem, że ten współczynnik istnieje i ma wartość skończoną.

Układem astatycznym nazywamy natomiast układ, w którym uchyb statyczny (tzn. uchyb ustalony przy wymuszeniu skokowym) jest równy zero. Jest to możliwe tylko wówczas, gdy k_0 jest nieskończone, tzn. $\lim_{s \rightarrow 0} G_0(s) = \infty$. To oznacza,

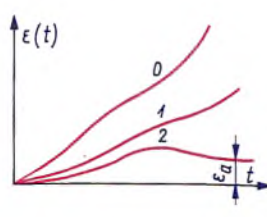
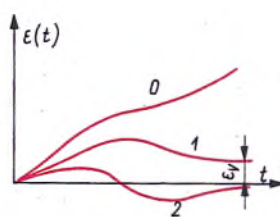
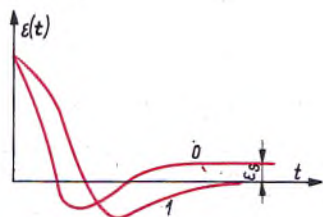
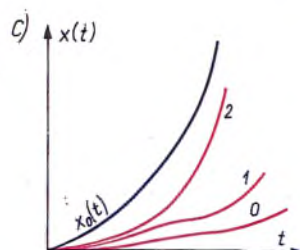
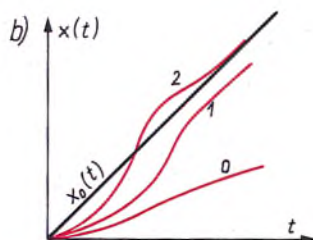
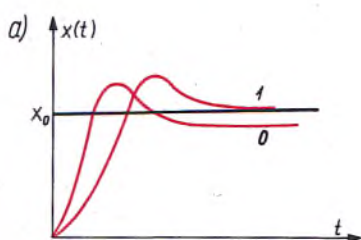
że transmitancja układu otwartego ma biegun w zerze, czyli układ otwarty zawiera człon całkujący w połączeniu szeregowym z dowolnymi członami o skończonym wzmocnieniu, a więc $G_0(s)$ ma postać $\frac{k_v}{s} G'_0(s)$, przy czym $G'_0(0) = 1$, k_v — współczynnik wzmocnienia prędkościowego.

Przy wymuszeniach zmieniających się liniowo, np. $x_0(t) = A_1 \cdot t$, czyli $X_0(s) = \frac{A_1}{s^2}$, układ regulacji statycznej nie może w ogóle działać, ponieważ uchyb ustalony jest nieskończony, natomiast układ astatyczny wykazuje tzw. **uchy-
b prędkościowy**

$$\varepsilon_v = \lim_{s \rightarrow 0} \frac{A_1}{s} \frac{1}{1 + G_0(s)} = \frac{A_1}{k_v}$$

Aby uchyb prędkościowy był równy zero, jest potrzebna cecha astatyzmu drugiego stopnia, tzn. transmitancja $G_0(s)$ powinna mieć biegun podwójny w zerze, a więc zawierać

•3.8• Ilustracja pojęcia astatyzmu — odpowiedzi układów regulacji: a) na skok x_0 ; b) na sygnał $x_0(t)$ liniowy; c) na sygnał $x_0(t)$ paraboliczny:
0 — układ statyczny; 1 — astatyczny; 2 — astatyczny drugiego stopnia



dwa całkowania $G_0(s) = \frac{k_a}{s^2} G'_0(s)$, $G'_0(0) = 1$, k_a — współczynnik wzmocnienia przyspieszeniowego. Łatwo jednak sprawdzić, że i układ astatyczny drugiego stopnia mógłby wykazywać uchyb ustalony, gdyby wymuszenie było paraboliczne — $x_0(t) = A_2 t^2$, $X_0(s) = \frac{2A_2}{s^3}$. Uchyb ustalony przy wymuszeniu parabolicznym nazywa się **uchybem przyspieszeniowym**, jego wartość wynosi $\varepsilon_a = \frac{2A_2}{k_a}$. Pojęcie stopnia astatyzmu można rozumieć ogólnie, choć praktycznie rozważa się tylko stopień zerowy (statyzm), pierwszy i drugi. Oczywiście układ astatyczny wyższego stopnia ma zerowe uchyby ustalone wszystkich „niższych” stopni. Zilustrowano to na rys. 3.8.

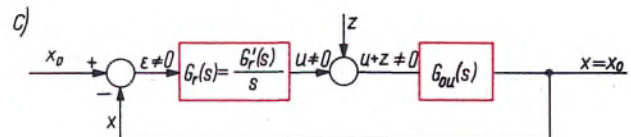
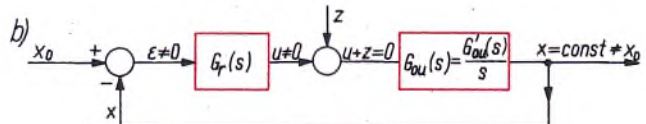
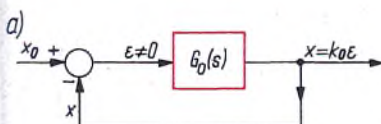
3.9. Czy astatyzm względem sygnału zadanego jest równoważny astatyzmowi względem zakłócenia?

Nie, w poprzednim pytaniu rozważono tylko najprostszy przypadek zakłócenia na wyjściu obiektu regulacji, ale ogólnie ta równoważność nie zachodzi. Jeśli zakłócenie wchodzi na wejściu obiektu regulacji, czyli $G_{oz}(s) = G_{ou}(s)$ i $G_z(s) = G(s)/G_r(s)$, to astatyzm względem zakłócenia wystąpi wtedy, gdy transmitancja regulatora $G_r(s)$ będzie miała biegun w zerze, gdyż tylko wtedy można otrzymać $\lim_{s \rightarrow 0} G_z(s) = 0$, biorąc pod uwagę, że zawsze obowiązuje zależność $\lim_{s \rightarrow 0} G(s) \neq 0$. Może więc się zdarzyć, że układ jest astatyczny

względem wartości zadanej, a nie jest astatyczny względem zakłócenia — jeśli transmitancja $G_r(s)$ nie zawiera całkowania, a zawiera je transmitancja obiektu $G_{ou}(s)$. Obiekt zawierający całkowanie często nazywa się obiektem astatycznym, a nie zawierający całkowania — statycznym.

Sens astatyzmu można dość prosto zinterpretować. Jeśli układ regulacji w ogóle nie zawiera członów całkujących, to między wszystkimi sygnałami w stanie ustalonym zachodzą zależności proporcjonalne ze współczynnikami proporcjonalności różnymi od zera. Element całkujący jest jedynym elementem, który może mieć zero na wejściu, a niezerowy

3.9. Zależności pomiędzy wartościami ustalonymi sygnałów w układzie regulacji: a) układ statyczny, transmitancja $G_0(s)$ nie zawiera całkowania; b) całkowanie w obiekcie, układ astatyczny względem wartości zadanej, statyczny względem zakłócenia; c) całkowanie w regulatorze, układ astatyczny względem wartości zadanej i zakłócenia



(stały) sygnał na swym wyjściu. Wobec tego struktura układu regulacji powinna być taka, aby uchyb regulacji był sygnałem wejściowym elementu całkującego. Na rysunku 3.9 przedstawiono kilka możliwych sytuacji. Z poprzednich rozważań i z tego rysunku wynika, że w układzie astatycznym względem zakłócenia — zakłócenie to powinno wchodzić za elementem całkującym (nie może być całkowane).

•3.10• W jaki sposób ocenia się właściwości układu regulacji w stanie ustalonym przy dowolnych wymuszeniach?

Omówione poprzednio pojęcia dotyczą tylko reakcji układu regulacji na wymuszenie (wartość zadana, zakłócenia) o postaci prostych funkcji potęgowych (lub wielomianowych). Zauważmy jednak, że przy wymuszeniach sinusoidalnych pojęcia te nie mają zastosowania, ponieważ stan ustalony przy wymuszeniach sinusoidalnych w przypadku układu liniowego oznacza, że wszystkie przebiegi mają postać sinusoidalną, różniąc się co najwyżej amplitudą i fazą. Najwygodniej wtedy operować transmitancjami widmowymi. W szczególności transmitancja widmowa uchybowa ma wtedy postać

$$G_e(j\omega) = \frac{1}{1 + G_0(j\omega)}$$

i jest nazywana **częstotliwościowym wskaźnikiem regulacji** $q(j\omega)$ (patrz pyt. 3.6), ponieważ pokazuje, jak silnie jest tłumiony uchyb pochodzący od sygnału wymuszającego (x_0 lub z) w układzie zamkniętym regulacji w stosunku do układu otwartego. Dla wymuszenia sinusoidalnego o danej pulsacji ω_0 i o amplitudzie A można więc od razu powiedzieć, że uchyb będzie także sinusoidalny — o pulsacji ω_0 i amplitudzie $A|q(j\omega_0)|$.

Inne wymuszenia okresowe można rozłożyć w szereg Fouriera. Składowe harmoniczne o pulsacjach $\omega_0, 2\omega_0, 3\omega_0, \dots$ są wówczas tłumione odpowiednio do wartości $|q(jn\omega_0)|$ przy $n = 1, 2, \dots$ W przypadku wymuszeń scharakteryzowanych widmem ciągłym — np. w przypadku zakłóceń o charakterze funkcji losowych — istotny staje się cały przebieg charakterystyki częstotliwościowej $|q(j\omega)|$. Działanie układu regulacji jest tym lepsze, w im szerszym pasmie częstotliwości moduł $|q(j\omega)|$ ma małe wartości — oznacza to bowiem odpowiednio małe wartości amplitud uchybu. Pasma regulacji, w którym jest spełniony warunek $|q(j\omega)| \approx 0$, jest jednak zawsze ograniczone, a więc przybliżoną eliminację uchybu można stosować tylko dla wymuszeń (x_0 lub z) oddziałujących w tym ograniczonym pasmie.

•3.11• Czy zadania dotyczące dokładności mają wpływ na jakość przebiegów przejściowych?

Tak, jest to podstawowy konflikt przy projektowaniu układów regulacji. Żądania dużej dokładności (zwłaszcza dotyczące nie tylko stanu ustalonego) zagrażają **stabilności** układu regulacji. Stabilność jest jednak cechą, której kategorię nie można naruszyć — niestabilność oznacza bowiem „najgorszą” możliwą jakość przebiegów przejściowych.

•3.12• W jaki sposób ocenia się stabilność układu regulacji?

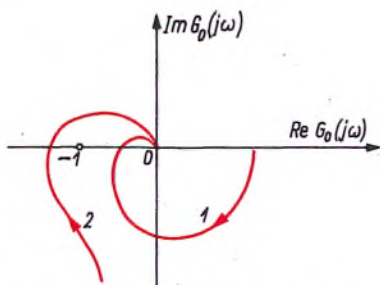
Zdecydowanie najwięcej korzyści daje ocena stabilności układu zamkniętego regulacji na podstawie właściwości układu otwartego. Sposób taki ma głębokie uzasadnienie praktyczne, ponieważ budując układ regulacji dobrze jest wiedzieć, czy po zamknięciu pętli sprzężenia zwrotnego pomiędzy obiektem a regulatorem układ będzie stabilny; właściwości układu otwartego są zazwyczaj znane (lub mogą być określone eksperymentalnie) wcześniej, zanim cały układ regulacji jest zrealizowany.

Ocena stabilności układu zamkniętego na podstawie układu otwartego opiera się na spostrzeżeniu, że w mianowniku wszystkich wyrażeń określających transmitancję układu zamkniętego regulacji występuje czynnik $1 + G_o(s)$. Przyrównując to wyrażenie do zera

$$1 + G_o(s) = 0$$

można je uważać za równanie charakterystyczne układu zamkniętego, tzn. wartości zmiennej s , dla których powyższe równanie jest spełnione, są biegunami układu zamkniętego. Należy stwierdzić, czy wszystkie te bieguny leżą w lewej półpłaszczyźnie zmiennej zespolonej; jeśli nie — układ zamknięty jest niestabilny. Problem ten można badać różnymi metodami. Można więc po prostu obliczyć wartości pierwiastków równania i sprawdzić, czy ich części rzeczywiste są mniejsze od zera; jest to postępowanie żmudne i nie pokazujące możliwych typów zachowania się układu zamkniętego przy zmianie parametrów układu. W stosunku do takiej metody bezpośredniej pewien postęp wnoszą tzw. kryteria algebraiczne badania stabilności, stanowiące pewne warunki nałożone na współczynniki równania charakterystycznego; spełnienie tych warunków gwarantuje ujemność części rzeczywistej pierwiastków. Tak więc kryteria algebraiczne zwalniają nas od obliczania wartości pierwiastków, ale też rozstrzygają tylko kwestię położenia tych pierwiastków w lewej lub prawej półpłaszczyźnie. W najprostszym przypadku równania charakterystycznego drugiego stopnia współczynniki tego równania muszą być dodatnie, aby pierwiastki miały ujemne części rzeczywiste (łatwo to sprawdzić na przykładzie równania kwadratowego). W przypadku równań wyższych stopni warunek dodatniości współczynników nadal obowiązuje, ale dochodzi jeszcze szereg bardziej skomplikowanych warunków. Kryteria algebraiczne nie będą dokładniej omówione.

Stwierdziłszy w poprzednich pytaniach, że w układzie re-



•3.13A• Ilustracja kryterium stabilności Nyquista

1 — charakterystyki amplitudowo-fazowe układu otwartego, który po zamknięciu będzie stabilny; 2 — który po zamknięciu będzie niestabilny

gulacji bardzo istotne znaczenie ma współczynnik wzmocnienia układu otwartego regulacji. Przyjmijmy więc dla wygody, że transmitancję wypadkową układu otwartego można przedstawić w postaci $G_0(s) = KG'_0(s)$, przy czym K jest liczbą dodatnią określającą wzmocnienie w pętli regulacji. Wtedy równanie charakterystyczne układu zamkniętego można zapisać jako

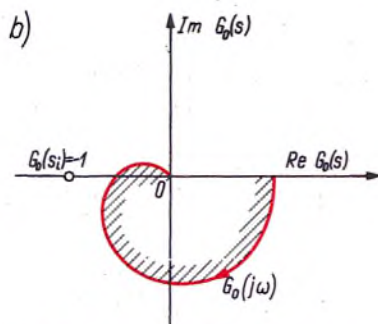
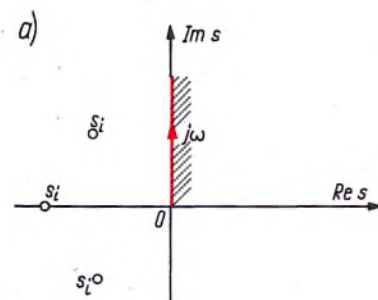
$$1 + KG'_0(s) = 0$$

$$\text{czyli } G'_0(s) = -\frac{1}{K}$$

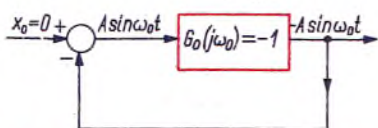
Zależność ta jest punktem wyjścia dla dwu metod oceny właściwości układu zamkniętego (w tym jego stabilności): metody Nyquista i metody linii pierwiastkowych.

•3.13• Na czym polega metoda Nyquista badania stabilności?

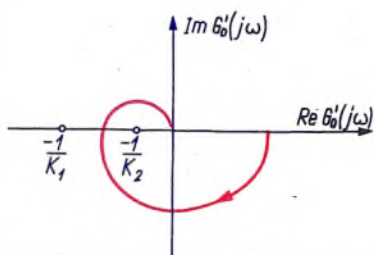
Metoda Nyquista jest metodą częstotliwościową i dzięki temu można ją sprawdzić doświadczalnie. W najprostszym i najczęściej spotykanym przypadku, kiedy układ otwarty jest stabilny, ocena stabilności układu zamkniętego wymaga tylko wykreślenia charakterystyki amplitudowo-fazowej układu **otwartego** we współrzędnych $\text{Re } G_0(j\omega)$, $\text{Im } G_0(j\omega)$, przy czym — co istotne — charakterystykę tę można określić eksperymentalnie. **Kryterium stabilności Nyquista** wówczas brzmi: układ zamknięty otrzymany z danego układu otwartego jest stabilny wtedy i tylko wtedy, gdy charakterystyka amplitudowo-fazowa nie **obejmuje** punktu $(-1, j0)$ (rys. 3.13A). Niezbyt ściśle, ale poglądowe uzasadnienie tego kryterium można oprzeć na spostrzeżeniu, że charakterystyka amplitudowo-fazowa stanowi odwzorowanie osi urojonej $j\omega$ płaszczyzny zmiennej zespolonej s przy $0 \leq \omega < \infty$, przy czym odwzorowanie określa funkcja $G_0(s)$. Odwzorowanie to, przy funkcji $G_0(s)$ wymiernej, zachowuje podstawowe właściwości odwzorowywanych obszarów, w szczególności lewa i prawa półpłaszczyzna zmiennej s przechodzi odpowiednio w obszary leżące po lewej i prawej stronie obrazu osi $j\omega$, czyli charakterystyki amplitudowo-fazowej. Ponieważ bieguny układu **zamkniętego** znajdują się wszystkie w punkcie $G_0(s) = -1$ (patrz poprzednie pytanie), więc jeśli układ zamknięty ma być stabilny, to bieguny te powinny leżeć w **lewej** półpłaszczyźnie zmiennej s , zatem po odwzorowaniu przez funkcję $G_0(s)$ obraz tych biegunów (czyli punkt -1) powinien także znaleźć się na **lewo** od obrazu osi $j\omega$, jak to sugeruje rys. 3.13B. Przekonującą interpretację otrzymamy przyjmując podejście częstotliwościowe i używając transmitancji widmowych. Warunek $G_0(j\omega) = -1$ oznacza, że istnieje taka pulsacja $\omega = \omega_0$, przy której sygnał sinusoidalny może się sam podtrzymać w pętli regulacji, skoro po przejściu przez układ otwarty i pętlę **ujemnego** sprzężenia zwrotnego pojawia się na wejściu układu otwartego z tą samą



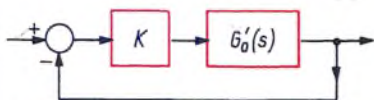
•3.13B• Odwzorowanie płaszczyzny zmiennej zespolonej s z płaszczyzną $G_0(s)$



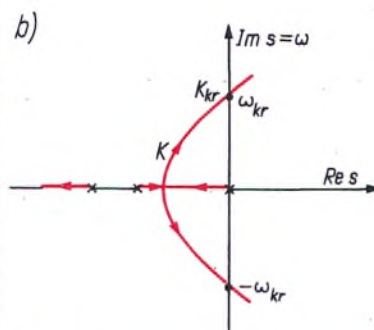
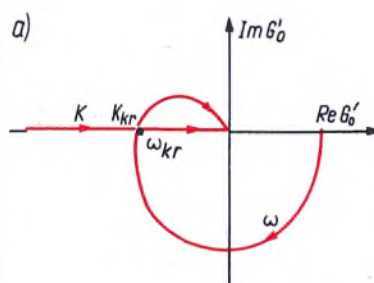
•3.13C• Interpretacja warunku $G_0(j\omega) = -1$ dla $\omega = \omega_0$



• 3.13D • Interpretacja kryterium Nyquista przy $G_0(j\omega) = KG'_0(j\omega)$; układ zamknięty stabilny przy $K = K_1$, niestabilny przy $K = K_2$



• 3.14A • Schemat blokowy układu regulacji stosowany przy badaniu linii pierwiastkowych



• 3.14B • Interpretacja krytycznego przypadku kryterium Nyquista: a) we współrzędnych charakterystyk amplitudowo-fazowych (przecięcie charakterystyki z osią rzeczywistą); b) we współrzędnych płaszczyzny zmiennej zespolonej (przecięcie linii pierwiastkowych z osią urojoną)

amplitudą i fazą (rys. 3.13C). Dla $\omega = \omega_0$ efektywne sprzężenie zwrotne staje się więc **dodatnie** — z racji przesunięcia fazy o π .

Zauważmy jeszcze, że przyjęcie $G_0(j\omega) = KG'_0(j\omega)$ powoduje wygodną modyfikację wykresu Nyquista: charakterystyka $G'_0(j\omega)$ nie powinna wtedy obejmować punktu $-\frac{1}{K}$ (rys. 3.13). Rysunki 3.13B i 3.13D są oczywiście równoważne, natomiast rozważanie zależności od współczynnika wzmocnienia K staje się wygodniejsze.

• 3.14 • Co to jest linia pierwiastkowa?

Linia pierwiastkowa jest to miejsce geometryczne (na płaszczyźnie zmiennej zespolonej) pierwiastków równania charakterystycznego **układu zamkniętego**, otrzymane przy uzmiennieniu jednego z parametrów (najczęściej współczynnika wzmocnienia) układu **otwartego**. Dla schematu blokowego układu regulacji przedstawionego na rys. 3.14A równanie charakterystyczne układu zamkniętego jest równoważne równaniu rozważanemu już w poprzednich pytaniach

$$G'_0(s) = -\frac{1}{K}$$

a zatem możliwe położenie pierwiastków jest określone przez **warunek argumentu**

$$\arg G'_0(s) = n\pi; \quad n \text{ nieparzyste dla } K > 0$$

oraz **warunek modułu**

$$|G'_0(s)| = \frac{1}{|K|}$$

Z warunku argumentu korzysta się dla określenia kształtu linii pierwiastkowej, z warunku modułu zaś dla jej „przeskalowania”, tzn. określenia położenia pierwiastków na linii pierwiastkowej przy konkretnych wartościach K . Określanie linii pierwiastkowych z definicji byłoby dość trudne, dlatego podaje się zwykle pewne reguły, umożliwiające szybkie oszacowanie kształtu wykresu linii pierwiastkowych. Są to następujące reguły:

1. Przy układzie rzędu n wykres ma n gałęzi odpowiadających przesuwaniu się pierwiastków przy $0 \leq K < \infty$.
2. Wykres jest symetryczny względem osi rzeczywistej płaszczyzny zmiennej zespolonej s .
3. Każdy punkt osi rzeczywistej należy do wykresu, jeśli na prawo od tego punktu znajduje się ogółem nieparzysta liczba zer i biegunów rzeczywistych transmitancji $G_0(s)$.
4. Przy K bliskim zeru n gałęzi wykresu zaczyna się w n biegunach transmitancji $G_0(s)$.
5. Przy $K \rightarrow \infty$ m gałęzi wykresu dąży do m zer transmitancji $G_0(s)$; pozostałych $n - m$ gałęzi dąży asymptotycznie do punktów leżących w nieskończoności.

6. Asymptoty gałęzi wymienionych w p. 5 przecinają się na osi rzeczywistej w tzw. środku ciężkości biegunów i zer, leżącym w punkcie $\sigma_0 = (\sum_{i=1}^n s_i - \sum_{j=1}^m s_{0j}) / (n-m)$, przy czym s_i oznaczają bieguny, a s_{0j} — zera transmitancji $G_0(s)$; ponadto asymptoty te tworzą z osią rzeczywistą kąt

$$\Psi_k = (\pi + k \cdot 2\pi) / (n-m); \quad k = 0, 1, \dots, (n-m-1)$$

Najbardziej oczywiste są reguły 4 i 5 — wynikają one bezpośrednio z równości $G'_0(s) = -\frac{1}{K}$ przy $K \rightarrow 0$ i $K \rightarrow \infty$.

Zwróćmy też uwagę, że wykres linii pierwiastkowej uzupełnia kryterium Nyquista. Niestabilność układu zamkniętego oznacza, że jego biegun (lub bieguny) przechodzą do prawej półpłaszczyzny. Przecięcie osi urojonej przez linie pierwiastkową odpowiada równości $G'_0(j\omega) = -\frac{1}{K}$, czyli krytycznemu

przypadkowi kryterium Nyquista, tzn. współrzędne $\omega = \omega_{kr}$ i $K = K_{kr}$ są na obydwu wykresach identyczne (rys. 3.14B). Wynika to z faktu, że o ile odwzorowaniem osi $j\omega$ przez funkcję $G'_0(j\omega)$ jest charakterystyka amplitudowo-fazowa, to linie pierwiastkowe są przez tę samą funkcję $G'_0(j\omega)$ odwzorowane w ujemną półoś rzeczywistą (zgodnie z warunkiem argumentu).

Wadą metody linii pierwiastkowych jest ograniczenie do analizy wpływu jednego tylko parametru (wzmocnienia). Metoda ta nie jest powszechnie stosowana, chociaż pomaga

3.14C. Przykłady linii pierwiastkowych:

a) $G_0 = K \frac{1}{s-s_1}$;

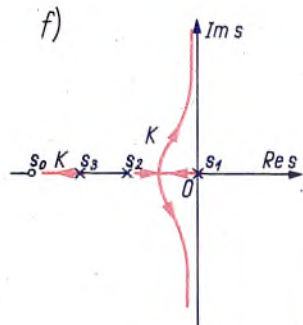
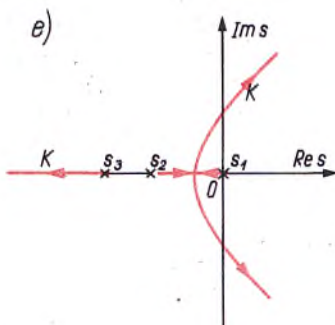
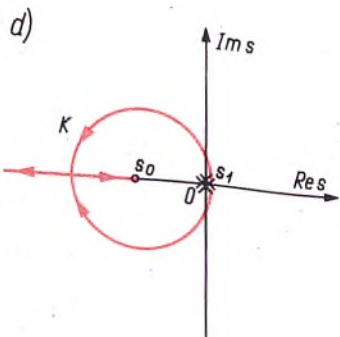
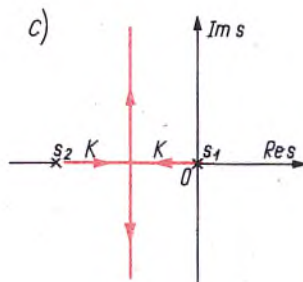
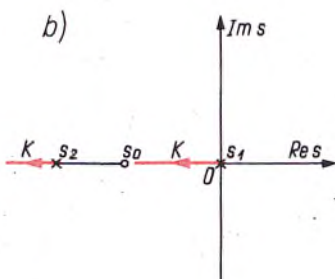
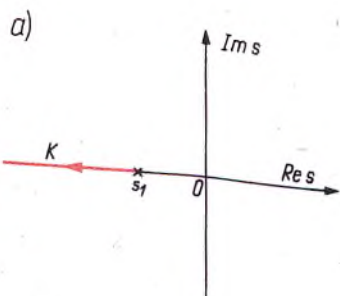
b) $G_0 = K \frac{s-s_0}{(s-s_1)(s-s_2)}$ (dla $s_1 = 0$);

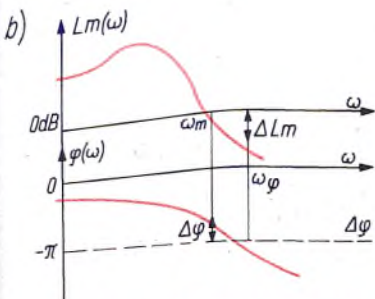
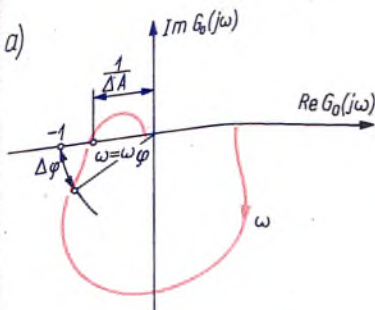
c) $G_0 = K \frac{1}{(s-s_1)(s-s_2)}$ (dla $s_1 = 0$);

d) $G_0 = K \frac{s-s_0}{(s-s_1)^2}$ (dla $s_1 = 0$);

e) $G_0 = K \frac{1}{(s-s_1)(s-s_2)(s-s_3)}$ (dla $s_1 = 0$);

f) $G_0 = K \frac{s-s_0}{(s-s_1)(s-s_2)(s-s_3)}$ (dla $s_1 = 0$);





3.15 Interpretacja zapasów stabilności: a) we współrzędnych amplitudowo-fazowych; b) we współrzędnych wykresów Bodego

$$\Delta Lm = 20 \lg \Delta A, \text{ zapas fazy } - \Delta \varphi$$

wyrobić sobie pewną intuicję co do wpływu sprzężenia zwrotnego na układ, a ponadto daje wiele informacji o właściwościach układu, ponieważ bieguny transmitancji (wartości własne układu) określają ściśle zachowanie się układu. Na rysunku 3.14C pokazano kilka charakterystycznych wykresów linii pierwiastkowych.

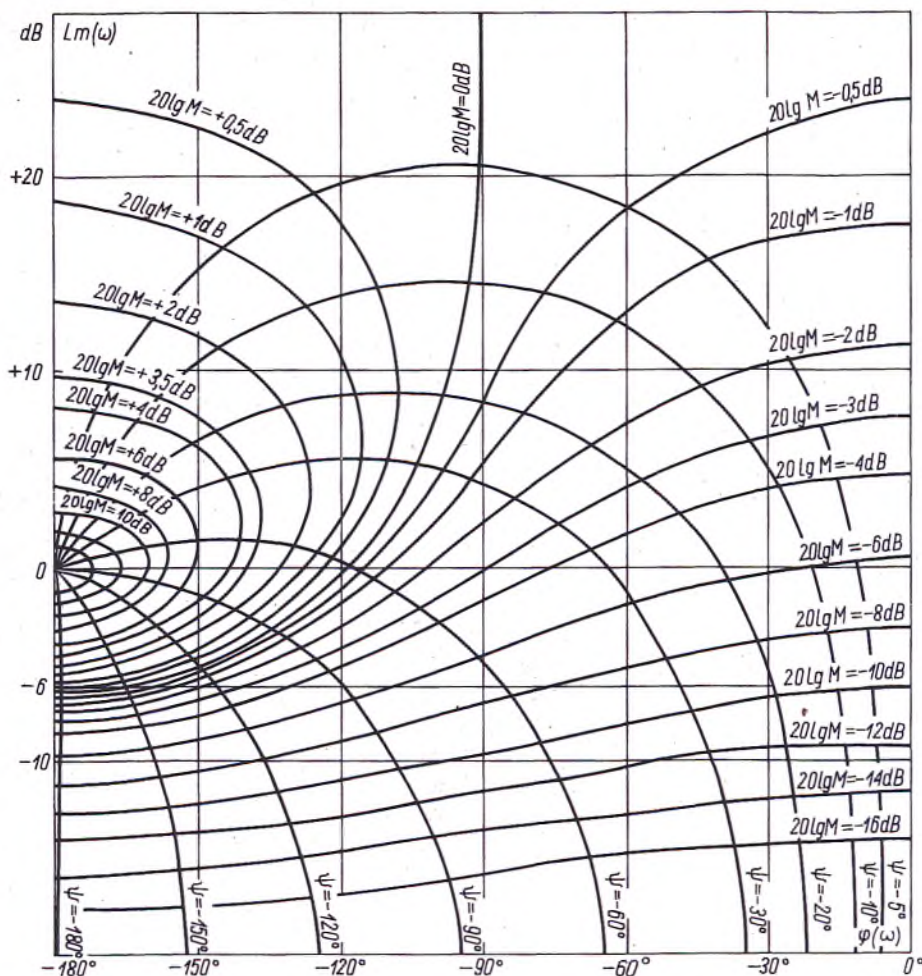
3.15 Co to jest zapas stabilności?

Jest to miara oddalenia charakterystyki $G_0(j\omega)$ od punktu krytycznego $(-1, j0)$. Oddalenie to ocenia się najwygodniej za pomocą **zapasu modułu i zapasu fazy**. Sens tych pojęć wyjaśnia rys. 3.15. Zapas modułu jest określony w punkcie $\omega = \omega_\varphi$ (częstotliwość odcięcia fazy), w którym przesunięcie fazowe jest równe $-\pi$; badamy wówczas, ile **razy** mógłby być wtedy większy moduł transmitancji, zanim osiągnąłby wartość 1, czyli 0 dB. Zapas fazy jest określony w punkcie $\omega = \omega_m$ (częstotliwość odcięcia modułu), w którym moduł transmitancji jest równy 1 (czyli 0 dB); badamy wówczas, o ile zwiększy się przesunięcie fazowe, zanim osiągnie wartość $-\pi$. Interpretacja zapasu modułu ΔLm i zapasu fazy $\Delta \varphi$ jest szczególnie wyraźna na wykresach Bodego (rys. 3.15b), gdzie obydwa zapasy są określone wprost przez rzędne wykresów charakterystyk $Lm(\omega)$ i $\varphi(\omega)$ w punktach ω_φ i ω_m . Na tychże charakterystykach kryterium Nyquista interpretuje się jako nierówność $\omega_m < \omega_\varphi$, tzn. stabilność układu zamkniętego wymaga, aby w układzie otwartym charakterystyka amplitudowa najpierw zmalała do 0 dB (do wzmocnienia 1), zanim przesunięcie fazowe osiągnie wartość $-\pi$.

3.16 Jakie cechy powinien mieć „dobry” układ regulacji?

Pojęcie „dobrego” układu regulacji jest względne, o czym wspomniano wcześniej. Spróbujmy jednak odpowiedzieć na postawione pytanie operując pojęciami wprowadzonymi w poprzednich pytaniach. Tak więc układ regulacji powinien zapewniać mały (najlepiej zerowy) uchyb ustalony — należy zatem zapewnić stopień astatyzmu odpowiedni dla typowych spodziewanych postaci wymuszeń. Sprawa dokładności w stanie ustalonym jest w tej sytuacji określona jednoznacznie: albo układ jest astatyczny i nie wykazuje uchybu ustalonego (np. przy wymuszeniach skokowych), albo jest to układ statyczny i charakteryzuje się tzw. współczynnikiem statyzmu równym $1/(1+k_0)$ — por. pyt. 3.8.

Ocena właściwości dynamicznych układu jest w dużym stopniu umowna. W większości przypadków uważa się za „dobre” przeregulowanie w odpowiedzi skokowej nie przekraczającej 20% przy możliwie małym czasie regulacji lub — co jest w przybliżeniu równoważne — moduł rezonansowy charakterystyki układu zamkniętego nie większy niż 1,5 (3,5 dB) przy możliwie szerokim pasmie roboczym. Dalsze równoważne (w przybliżeniu) warunki sprowadzają się do żądania (równocześnie) zapasu modułu nie mniej niż 6 dB



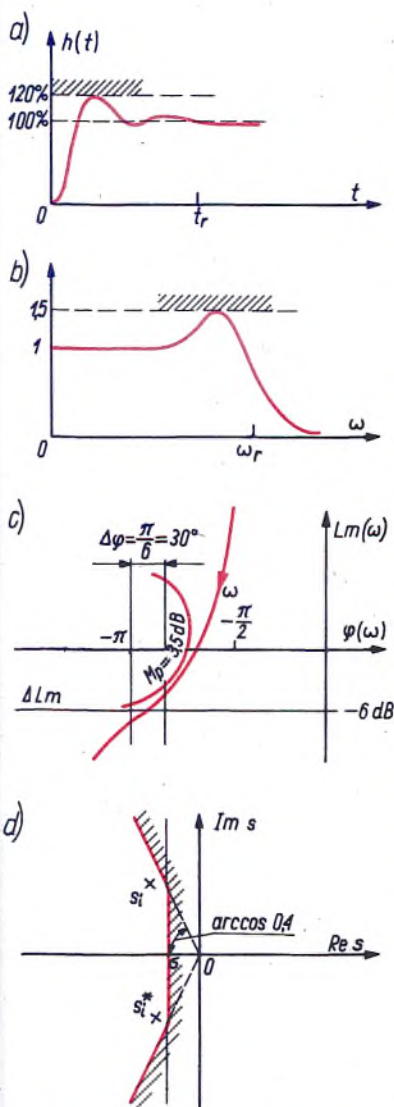
•3.16A• Karta Nicholasa

$Lm(\omega)$ i $\varphi(\omega)$ — współrzędne charakterystyk układu otwartego; M i Ψ — współrzędne charakterystyk układu zamkniętego

i zapasu fazy nie mniej niż $\frac{\pi}{6}$ (30°) — oczywiście również przy możliwie dużych częstotliwościach granicznych.

Mówiąc o zapasach stabilności i o ocenie właściwości układu zamkniętego na podstawie właściwości układu otwartego, należy wspomnieć o tzw. **karcie Nicholasa** (rys. 3.16A). Jest to wykres we współrzędnych moduł logarytmiczny-argument, na który naniesiono siatkę linii stałych wartości modułu transmitancji układu zamkniętego (M) i fazy układu zamkniętego Ψ . Rysując na karcie Nicholasa charakterystykę częstotliwościową układu otwartego (ω jest parametrem wykresu), można ocenić właściwości układu zamkniętego nie tylko pośrednio za pomocą zapasów modułu i fazy, ale od razu także widać, czy wartość M jest bliska 1 dla małych ω i jaka jest wartość modułu rezonansowego M_p .

W odniesieniu do położenia pierwiastków równania charakterystycznego, wymienione warunki można formułować inaczej, wykorzystując spostrzeżenie, że o charakterze oscyla-



·3.16B· Interpretacja cech „dobrego” układu regulacji: a) przeregulowanie w odpowiedzi skokowej mniejsze od 20% przy małym czasie regulacji; b) moduł rezonansowy mniejszy niż 1,5 przy szerokim pasmie regulacji; c) przebieg charakterystyk układu otwartego na karcie Nicholasa — zachowane zapasy amplitudy 6 dB i fazy 30° , nieprzekroczenie pary dominujących pierwiastków zespolonych — tłumienie bezwzględne σ możliwe duże, tłumienie względne większe niż 0,4

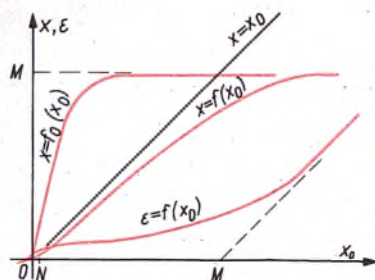
cyjnym odpowiedzi decyduje zwykle dominująca para pierwiastków zespolonych, najbliższa osi urojonej. W tej sytuacji warunek równowagi sprowadza się do określenia dopuszczalnego współczynnika tłumienia względnego nie mniejszego niż 0,4 przy możliwie dużym współczynniku tłumienia bezwzględnego (patrz pyt. 2.53). Na rysunku 3.16B zestawiono sposoby oceny jakości przebiegu przejściowego w różnych współrzędnych. Należy podkreślić, że podane oceny mają charakter orientacyjny.

·3.17· Jak można zinterpretować pojęcie astatyzmu w przypadku układu z nieliniowością statyki?

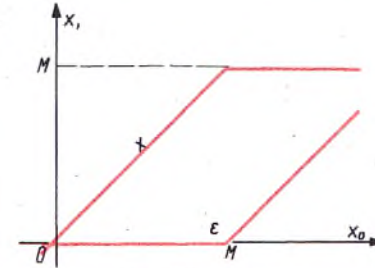
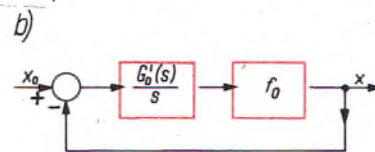
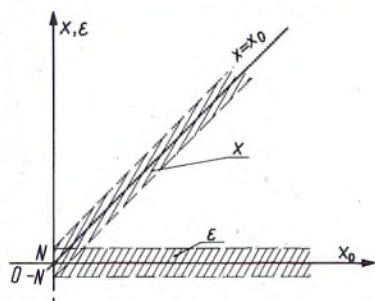
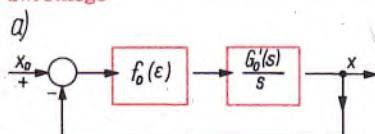
Najłatwiej ocenić sytuację odpowiadającą statyzmowi. Jeśli bowiem układ otwarty ma charakterystykę statyczną nieliniową, to można łatwo skonstruować charakterystykę statyczną układu zamkniętego, zgodnie z zasadami omówionymi w pyt. 2.33. Na rysunku 3.17A przedstawiono przykłady dość typowych kształtów charakterystyk statycznych układu otwartego i układu zamkniętego. Widać, że w tych zakresach charakterystyki układu otwartego, w których wykazuje ona „zerowe wzmocnienie” (a więc w obszarze strefy nieczułości lub w obszarze nasycenia), nie można dostrzec wpływu regulacji — uchyb zmienia się dokładnie tak samo jak wartość zadana. Efektów podobnych do astatyzmu można się spodziewać w przypadku występowania całkowania, jednakże wtedy dość trudno mówić o charakterystyce statycznej, gdyż nie jest ona określona. Jeśli jednak w strukturze fizycznej układu można oddzielić człony statyczne nieliniowe od członów dynamicznych liniowych zawierających całkowanie, to sytuacja nadaje się do analizy. W celu oceny właściwości układu w stanie ustalonym można przyjąć, że część liniowa zawierająca całkowanie jest równoważna elementowi o wzmocnieniu statycznym nieskończonym. Okaże się wówczas, że charakterystyki statyczne układu zamkniętego są różne w zależności od usytuowania członów. Na rysunku 3.17B przedstawiono dwie skrajne możliwości dla przypadku charakterystyki statycznej części nieliniowej jak na poprzednim rysunku. W przypadku (a) dominuje efekt strefy nieczułości: uchyb ustalony można zlikwidować co najwyżej z dokładnością do strefy nieczułości. W przypadku (b) dominuje efekt nasycenia: po osiągnięciu nasycenia układ regulacji, dotąd astatyczny, przestaje w ogóle działać poprawnie. Jednak w obu przypadkach uchyb ustalony, chociaż może występować, nie zależy ani od wartości zadanej, ani od wzmocnienia układu otwartego.

·3.18· W jaki sposób można zastosować częstotliwościowe kryteria badania stabilności do układów nieliniowych?

Jak powiedziano wcześniej w pyt. 3.11 do 3.13, w metodach częstotliwościowych bada się stabilność układu zamkniętego na podstawie charakterystyk częstotliwościowych układu otwartego, przy czym rolę zasadniczą pełni równość



•3.17A• Typowy kształt charakterystyk statycznych układu otwartego $f_0(x_0)$ i układu zamkniętego nie zawierającego całkowania w pętli sprzężenia zwrotnego



•3.17B• Wpływ obecności całkowania w pętli regulacji na charakterystyki statyczne układu zamkniętego: a) człon nieliniowy f_0 o charakterystyce jak na rys. 3.17A przed członem liniowym; b) człon nieliniowy za członem liniowym

$G'_0(j\omega) = -1/K$. W wielu przypadkach można w układzie otwartym oddzielić część liniową dynamiczną o transmittancji widmowej $G'_0(j\omega)$ i wzmacnieniu jednostkowym i część nieliniową statyczną. Jeśli przy tym potrafimy określić dla danej charakterystyki statycznej pewien „zastępczy punkt $-1/K$ ”, to możemy sądzić, że rozumowanie związane z kryterium Nyquista pozostanie w mocy. Rzeczywiście w pewnych przypadkach jest to możliwe; za najbardziej przydatne można uznać metodę funkcji opisującej oraz kryterium Kudrewicza-Cypkina.

•3.19• Na czym polega metoda funkcji opisującej?

Metoda funkcji opisującej wykorzystuje zasadę tzw. **linearyzacji harmonicznej**, której istotę wyjaśnia rys. 3.19A. Jeśli w układzie występują drgania nietłumione lub słabo tłumione, to będą one miały kształt w przybliżeniu sinusoidalny pod warunkiem, że część liniowa $G'_0(j\omega)$ ma właściwości silnie filtrujące (a więc charakterystyka modułu $|G'_0(j\omega)|$ silnie maleje przy wzroście ω). W tej sytuacji można uznać, że o stabilności będzie w przybliżeniu decydował „bilans” harmonicznej podstawowej przebiegów. Stąd powstała idea, aby część nieliniową opisać za pomocą „wzmocnienia” tylko harmonicznej podstawowej, obliczonej na podstawie rozkładu w szereg Fouriera przebiegu za członem nieliniowym (przebieg ten z pewnością nie jest sinusoidalny) przy założeniu, że po przejściu przez część liniową otrzymuje się na wejściu członu nieliniowego przebieg sinusoidalny. Jeśli charakterystykę statyczną członu nieliniowego można opisać zależnością $u = f(\varepsilon)$, to przy $\varepsilon = A \sin \omega t$ można ten człon nieliniowy scharakteryzować za pomocą funkcji opisującej

$$K(A) = g + jb$$

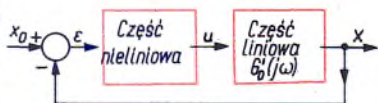
będącej liczbą zespoloną, przy czym

$$g(A) = \frac{1}{\pi A} \int_0^{2\pi} f(A \sin \omega t) \sin \omega t d\omega t$$

$$b(A) = \frac{1}{\pi A} \int_0^{2\pi} f(A \sin \omega t) \cos \omega t d\omega t$$

W prostszych przypadkach $K(A)$ zależy tylko od amplitudy A (nie zależy od ω), ponadto przy jednoznacznych charakterystykach $f(\varepsilon)$ część urojona $b(A)$ jest równa zero, tzn. nieliniowość nie powoduje przesunięcia fazy. Na rysunku 3.19B podano kilka przykładów funkcji opisujących. Jeśli w układzie zamkniętym mają powstać drgania o częstotliwości podstawowej ω_r i amplitudzie A_r , to powinien być spełniony warunek

$$G'(j\omega_r) = -\frac{1}{K(A_r)}$$



• 3.19A • Ilustracja zasady linearyzacji harmonicznej:

ε — przebieg sinusoidalny, u — przebieg odkształcony, zawierający wyższe harmoniczne, x — przebieg sinusoidalny (w przybliżeniu, po odfiltrowaniu wyższych harmonicznych)

• 3.19B • Przykłady funkcji opisujących:
a) element ze strefą nieczułości (funkcja opisująca rzeczywista); b) element z nasyceniem (funkcja opisująca rzeczywista); c) przekładnik idealny (funkcja opisująca rzeczywista); d) przekładnik idealny z histerezą (funkcja opisująca zespolona — podano moduł i argument)

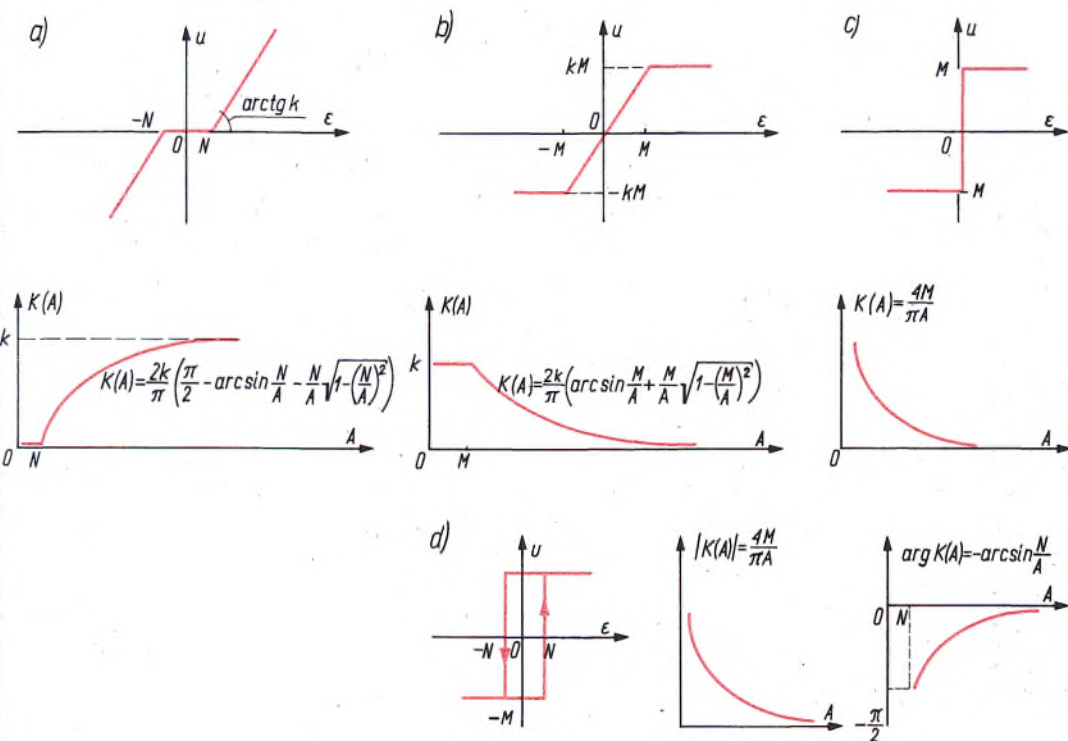
A więc, jeśli na płaszczyźnie zmiennej zespolonej wykreślimy charakterystykę amplitudowo-fazową $G'(j\omega)$ oraz funkcję opisującą (a raczej jej odwrotność z minusem) $-1/K(A)$, to przecięcie tych dwu krzywych świadczy o spełnieniu warunku powstania drgań nieliniowych. Na rysunku 3.19C przedstawiono kilka możliwych sytuacji; przy kilku przecięciach krzywych zwykle tylko jedno z nich odpowiada drganiom stabilnym (cyklowi granicznemu stabilnemu), co można dość prosto zbadać rozważając punkty w pobliżu przecięcia (rys. 3.19C (c)).

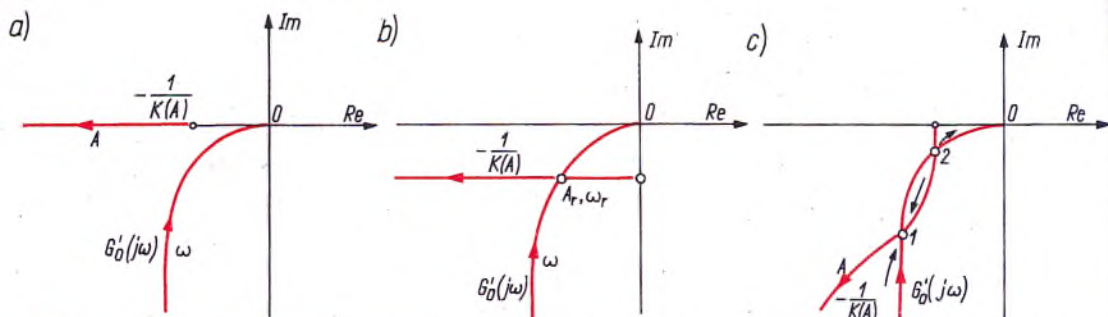
Należy wyraźnie podkreślić, że metoda funkcji opisującej jest metodą **przybliżoną**, a więc nie stanowi żadnych warunków koniecznych ani dostatecznych stabilności. Jeśli wykres $G_0'(j\omega)$ i $-1/K(A)$ wyraźnie się przecinają (lub nie przecinają), to można uznać, że świadczy to o niestabilności (lub stabilności) układu zamkniętego. Jeśli jednak wykresy leżą blisko siebie albo przecinają się „nieostro”, to właściwie nic nie można na tej podstawie powiedzieć o stabilności.

Wykresy $G_0'(j\omega)$ i $-1/K(A)$ można sporządzić również w układzie współrzędnych moduł logarytmiczny-argument.

• 3.20 • Na czym polega kryterium Kudrewicza-Cypkina?

Kryterium Kudrewicza-Cypkina dotyczy tylko pewnej klasy układów z nieliniowością statyki, mianowicie takiej, w której charakterystyka członu nieliniowego $f(\varepsilon)$ (rys. 3.20A) —





•3.19C. Zastosowanie funkcji opisującej do badania stabilności: a) układ stabilny (część liniowa typu całkowania z inercją, nieliniowość typu ograniczenia); b) układ zamknięty niestabilny przy częstotliwości ω_r i amplitudzie A_r (część liniowa jak poprzednio, nieliniowość typu przekaźnika z histerezą); c) dwa przecięcia wykresów $G'_0(j\omega)$ i $-\frac{1}{K(A)}$ (część liniowa jak

poprzednio, nieliniowość typu luzu mechanicznego); drgania stabilne występują w punkcie 1, gdzie zwiększenie ω powoduje zmniejszenie A i na odwrót, zaś w punkcie 2 drgania są niestabilne — co zaznaczono strzałkami

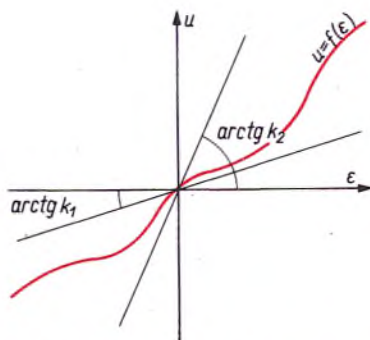
leży całkowicie pomiędzy dwiema prostymi o nachyleniach k_1 i k_2 , a część liniowa $G'_0(s)$ jest stabilna. Okazuje się wówczas, że rolę punktu krytycznego $-1/K$ pełni koło o środku na osi liczb rzeczywistych i brzegu przecinającym tę oś w punktach $-1/k_1$ i $-1/k_2$. Kryterium orzeka, że warunkiem **dostatecznym** stabilności asymptotycznej układu zamkniętego jest, aby charakterystyka amplitudowo-fazowa $G'_0(j\omega)$ nie przecinała ani nie obejmowała powyższego koła (podobnie jak to było w przypadku liniowym z punktem $-1/K$ w kryterium Nyquista). Na rysunku 3.20B pokazano kilka przykładów zastosowania omawianego kryterium. Nachylenia k_1 lub k_2 mogą być zerowe lub nieskończone — koło zamienia się wówczas w półpłaszczyznę. Zwracamy uwagę na dostateczność kryterium — a więc niemożność wyciągania ścisłych wniosków z przecięcia lub objęcia koła przez charakterystykę $G'_0(j\omega)$.

•3.21. W jaki sposób bada się stabilność, jeśli nie można zastosować żadnej z omówionych metod?

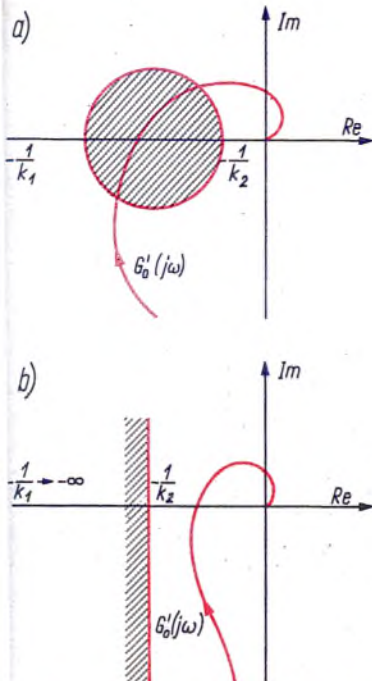
Jeśli układ jest silnie nieliniowy, lecz rząd układu nie jest wysoki, to najlepszą (ściśłą) metodą jest badanie trajektorii stanu, a zwłaszcza trajektorii fazowych. Jest to często metoda dość żmudna, jednak daje wyniki całkowicie jednoznaczne. Metoda ta ma szczególnie cenne zastosowanie do analizy układów nieliniowych, których nie można w żaden sposób linearyzować, w szczególności do układów przekaźnikowych, łącznie z przypadkiem występowania histerezy.

•3.22. Jakie etapy można wyróżnić w procesie projektowania układów regulacji?

Przy projektowaniu układów regulacji wstępnym etapem jest dokładne sprecyzowanie wymagań przyszłego użytkownika układu. Obiekt regulacji należy w zasadzie uznać za dany (choć w niektórych przypadkach procesy projektowania i budowy obiektu prowadzi się jednocześnie i łącznie z układami automatyki). Bardzo istotny jest wybór odpowiednich wejść sterujących i wyjść pomiarowych obiektu. Istota zadania musi być sformułowana w postaci wymagań dokładności i jakości przy jednoznacznym podaniu typowych (lub najgorszych spodziewanych) wymuszeń. Następnym ważnym



•3.20A. Charakterystyka czonu nieliniowego leżąca całkowicie między dwiema prostymi o nachyleniach k_1 i k_2



·3.20B· Przykłady zastosowania kryterium Kudrewicza-Cypkina: a) przecięcie koła przez charakterystykę $G'_0(j\omega)$ — wynik nieokreślony; b) układ zamknięty stabilny

etapem jest dobór oprzyrządowania sterowniczego, gwarantującego możliwość oddziaływania na obiekt w odpowiednich zakresach zmian sygnałów i przy odpowiednich mocach, a także oprzyrządowania pomiarowego zapewniającego rzetelną kontrolę wybranych zmiennych obiektu; na razie tym etapem nie będziemy się zajmować — zagadnienia techniki regulacji automatycznej omawiane są w rozdz. 4. Kolejnym bardzo ważnym etapem jest etap identyfikacji właściwości obiektu regulacji — z włączeniem elementów sterowniczych i pomiarowych. Podstawą tego etapu powinny być doświadczenia, w wyniku których uzyskamy model obiektu sformułowany w sposób ułatwiający zastosowanie go do celów regulacji, ale możliwie prosty. Dalsze etapy mają charakter bardziej uniwersalny; obejmują syntezę układu regulacji rozumianą jako określenie pożądanych zależności między sygnałami w układzie, zapewniających spełnienie postawionych wymagań, następnie zaś wybór struktury i określenie parametrów członów wprowadzanych do układu w charakterze elementów korekcyjnych. Uniwersalność tych etapów należy rozumieć przede wszystkim jako ich abstrakcyjność — operowanie opisem i sygnałami z pominięciem ich sensu fizycznego. W tym stadium projektant układu może jednak wykazać najwięcej inwencji, oczywiście popartej ugruntowaną wiedzą. Końcowy etap procesu projektowania ponownie nawiązuje do realiów fizycznych i polega na określeniu i sprawdzeniu zakresów zmian sygnałów w poszczególnych punktach układu, w szczególności sprawdzeniu, czy sygnały nie będą wykraczać poza strefę warunków znamionowych lub poza strefę liniowości. Etapy „techniczne” procesu projektowania wykazują istotne ograniczenie — najczęściej polegają na wyborze pewnego wariantu spośród dostępnego zestawu członów i urządzeń regulacyjnych.

·3.23· W jaki sposób identyfikuje się właściwości obiektu regulacji?

Identyfikację właściwości obiektu regulacji dzieli się zwykle na dwie części — identyfikację właściwości części statycznej i dynamicznej obiektu. Określenie charakterystyki statycznej w interesującym zakresie sygnałów pozwala ocenić możliwość przyjęcia modelu liniowego i ewentualnie określenie zakresu, w jakim taki model może obowiązywać. Identyfikacji części dynamicznej, jeśli wstępnie można przyjąć sensowność modelu liniowego, można dokonać za pomocą dwu podstawowych grup metod — czasowych i częstotliwościowych. Metody czasowe sprowadzają się do określenia charakterystyk czasowych obiektu, np. odpowiedzi na skok. Metody częstotliwościowe polegają na pomiarze charakterystyk częstotliwościowych. Pomiar taki jest często trudny lub nawet niemożliwy, jednak należy pamiętać, że charakterystyki częstotliwościowe zawierają znacznie więcej informacji o obiekcie niż charakterystyki czasowe. Jeśli więc tylko jest to wykonalne, należy dążyć do określenia charak-

terystyk częstotliwościowych, licząc się z dużym nakładem pracy z tym związanym oraz z koniecznością dysponowania specjalizowaną aparaturą pomiarową. Wymienione metody mają charakter czynny — wymagają wprowadzenia określonych sygnałów próbnych (skoków, sinusoid). W niektórych przypadkach takie postępowanie jest niedopuszczalne, pozostaje wtedy tylko metoda bierna, sprowadzająca się do obserwacji (rejestracji) sygnałów na wejściu i wyjściu obiektu a następnie do określenia zależności pomiędzy tymi sygnałami za pomocą metod statystycznych.

•3.24• Czy istnieją ogólne metody projektowania układów regulacji?

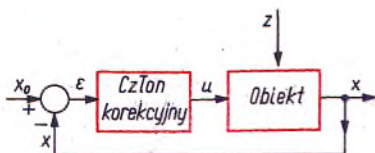
W odniesieniu do zadań nadążania lub kompensacji zakłóceń, gdy zadowalające wyniki daje przyjęcie modelu liniowego, za metodę ogólną można uznać zasadę korekcji charakterystyk częstotliwościowych, chociaż i ta zasada ulega modyfikacji w poszczególnych przypadkach, w zależności od spodziewanej postaci typowych wymuszeń. Natomiast zadanie przestawiania można rozwiązać zadowalająco tylko w układzie nieliniowym, wobec czego, jeśli tylko możliwe, należy dążyć do stosowania metody płaszczyzny fazowej.

•3.25• Co to jest korekcja charakterystyk częstotliwościowych?

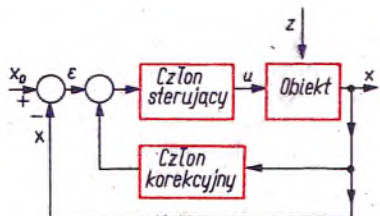
Jeśli jest dany obiekt regulacji i jego charakterystyki częstotliwościowe, to uznanie, że wystarczy wprowadzić sprzężenie zwrotne, dałoby wyniki bardzo dalekie od doskonałości. W przeważającej większości przypadków niezbędne jest umieszczenie regulatora w pętli sprzężenia zwrotnego po to, aby wypadkowe właściwości układu otwartego regulator-obiekt dawały pożądane cechy układu zamkniętego. Zasadniczą kwestię stanowi odpowiednie ukształtowanie charakterystyk częstotliwościowych układu otwartego, umożliwiające spełnienie postawionych warunków. Zabieg kształtowania wypadkowych charakterystyk układu otwartego przez regulator nazywa się korekcją. Istnieje kilka sposobów i kilka rodzajów korekcji. Najczęściej stosuje się **korekcję szeregową** (rys. 3.25A) oraz **korekcję przez sprzężenie zwrotne** (dodatkowe, oprócz sprzężenia głównego wytwarzającego sygnał uchybu regulacji — rys. 3.25B). Spośród stosowanych rodzajów korekcji można wyróżnić trzy główne działania: proporcjonalne (P), całkowe (I), różniczkowe (D) oraz ich kombinacje — proporcjonalno-całkową (PI), proporcjonalno-różniczkową (PD), proporcjonalno-całkowo-różniczkową (PID). Są to na tyle standardowe rodzaje korekcji, że obecnie dominującym typem regulatorów uniwersalnych, realizujących zadania korekcji dynamicznej, są **regulatory PID**.

•3.26• Co to jest regulator PID?

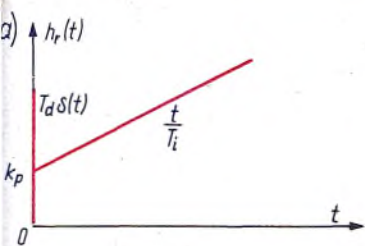
Regulator PID jest to korektor proporcjonalno-całkująco-



•3.25A• Korekcja szeregową

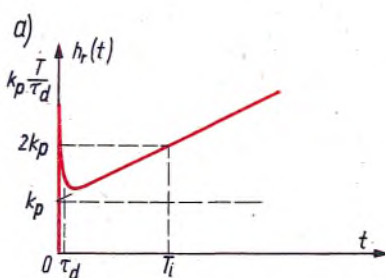


•3.25B• Korekcja przez dodatkowe sprzężenie zwrotne



3.26A· Charakterystyki idealnego regulatora PID: a) charakterystyka skokowa; b) charakterystyka częstotliwościowa

3.26B· Charakterystyki rzeczywistego regulatora PID: a) charakterystyka skokowa; b) charakterystyka częstotliwościowa



różniczkujący, tzn. realizujący w przypadku idealnym następującą zależność u od ε :

$$u(t) = k_p \left[\varepsilon(t) + \frac{1}{T_i} \int_0^t \varepsilon(\tau) d\tau + T_d \frac{d\varepsilon(t)}{dt} \right]$$

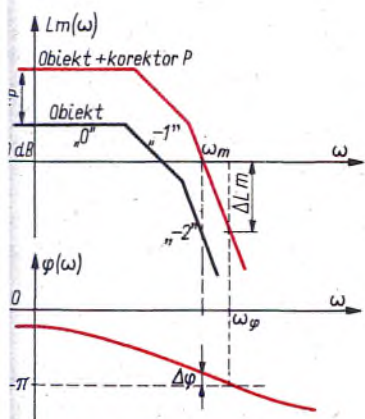
Odpowiada to transmitancji

$$G_r(s) = \frac{U(s)}{E(s)} = k_p \left(1 + \frac{1}{sT_i} + sT_d \right)$$

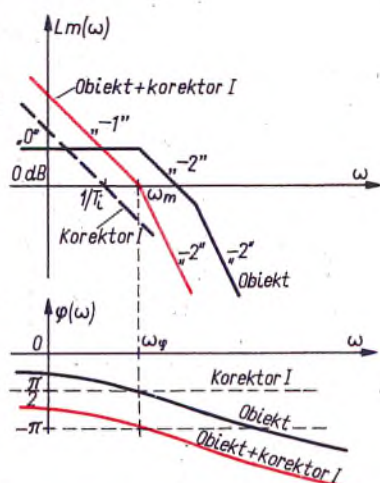
Parametr k_p jest współczynnikiem wzmocnienia proporcjonalnego regulatora, T_i — stałą całkowania, T_d — stałą różniczkowania. Na rysunku 3.26A pokazano charakterystykę skokową oraz charakterystyki częstotliwościowe regulatora PID w przypadku idealnym. Odpowiedź regulatora na dowolną postać sygnału ε jest ograniczona zakresem dopuszczalnych zmian sterowania u , wobec tego część D regulatora ma w rzeczywistości transmitancję $sT_d/(1+s\tau_d)$, przy czym τ_d jest zwykle znacznie mniejsze (co najmniej 10-krotnie) od T_d . Charakterystyki rzeczywistego regulatora PID przedstawiono na rys. 3.26B.

3.27· Na czym polega działanie P?

Działanie P (proporcjonalność) oznacza wzmocnienie uchybu. Działanie to zapewnia realizację elementarnego celu regulacji tj. zmniejszenie uchybu regulacji. Działanie P jest jednak działaniem dość „niebezpiecznym”, ponieważ łatwo prowadzi do niestabilności układu regulacji. Zinterpretowano to na rys. 3.27, przedstawiającym typowe charakterystyki częstotliwościowe obiektu regulacji oraz wypadkowego ukła-

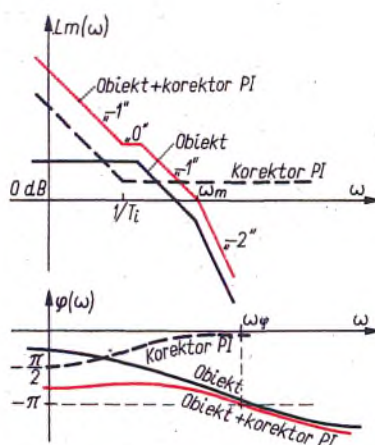


3.27· Idea korekcji P na charakterystykach częstotliwościowych (charakterystyka fazowa nie ulega zmianie)



•3.28A• Idea korekcji I na charakterystykach częstotliwościowych

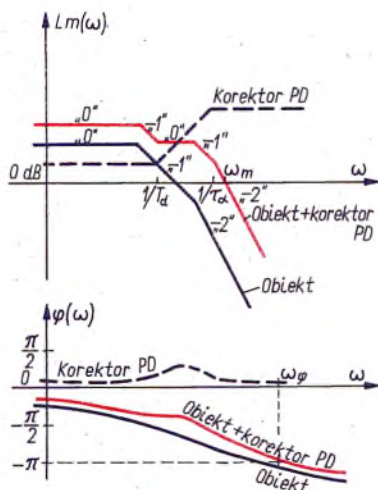
•3.28B• Idea korekcji PI na charakterystykach częstotliwościowych



du otwartego (korektor-obiekt) Widać, że przy wzroście wzmocnienia k_p zapasy stabilności szybko maleją, dając w zamian tylko pewne poszerzenie pasma roboczego i zmniejszenie współczynnika statyzmu $1/(1+k_0)$, jeśli obiekt regulacji był statyczny.

•3.28• Na czym polega działanie I?

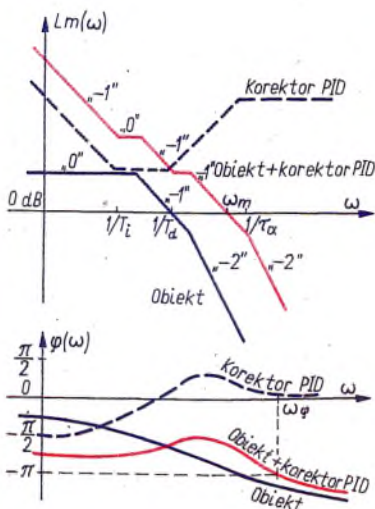
Działanie I (korekcja całkowa) jest to korekcja w zakresie małych częstotliwości, mająca za zadanie wprowadzić efekt astatyzmu. Samo działanie I jest jednak rzadko stosowane ze względu na zasadniczą wadę: co prawda w zakresie małych częstotliwości wzmocnienie układu otwartego jest wtedy bardzo duże, to jednak dla większych częstotliwości następuje bardzo duże zmniejszenie wzmocnienia i w wyniku tego ograniczenie szerokości pasma (rys. 3.28A). Nie bez znaczenia jest również przesunięcie fazy wprowadzane przez działanie I, równe $-\frac{\pi}{2}$, i powodujące pogorszenie warunków stabilności. Znacznie lepszy efekt daje działanie PI, które ma zalety działania I — wprowadza astatyzm — nie mając jego wad, ponieważ dla większych częstotliwości korektor PI zachowuje się tak jak P: nie wprowadza przesunięcia fazy i nie ogranicza pasma. Na rysunku 3.28B pokazano przykład korekcji PI. Warto zauważyć, że samo działanie I wystarczy do osiągnięcia elementarnego celu regulacji — likwidacji uchybu (przynajmniej ustalonego). Zauważmy też, że zmniejszanie wartości parametru T_i odpowiada zwiększaniu „wzmocnienia całkowania”.



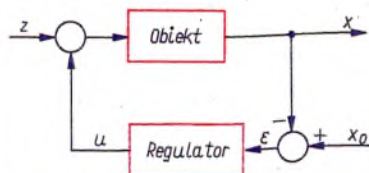
•3.29• Idea korekcji PD na charakterystykach częstotliwościowych

•3.29• Na czym polega działanie D?

Należy stwierdzić od razu, że działanie D nie zapewnia likwidacji uchybu — ma więc sens wyłącznie korekcyjny i wobec tego nigdy nie występuje samodzielnie, lecz wraz z działaniem P. Korektor PD jest to korektor dla zakresu większych częstotliwości. Zauważmy, że już przejście od działania



• 3.30• Idea korekcji PID na charakterystykach częstotliwościowych



• 3.31• Schemat blokowy układu regulacji przemysłowej (stałowartościowej)

I do PI jest równoważne wprowadzeniu korekcji PD do korektora I, ponieważ $\frac{1}{T_i s} (1 + T_d s) = \frac{T_d}{T_i} \left(\frac{1}{T_d s} + 1 \right)$. Korektor PD w ogólnym przypadku powoduje zwiększenie zapasu stabilności i wobec tego umożliwia rozszerzenie pasma regulacji przez dodatkowe zwiększenie wzmocnienia proporcjonalnego. Na rysunku 3.29 zilustrowano działanie korekcji PD. Korekcja PD nie wpływa na właściwości układu w stanie ustalonym.

• 3.30• Czy korekcja PID daje dobre rezultaty?

Tak, działanie PID przy całej swojej prostocie wykazuje bardzo cenne właściwości korekcyjne i w większości przypadków można dzięki niemu osiągnąć zadowalającą jakość regulacji. Działanie PID łączy zalety wszystkich trzech działań elementarnych: zapewnia astatyzm (I) przy rozszerzonym pasmie regulacji z zachowaniem zapasu stabilności (P, D). Na rysunku 3.30 przedstawiono wpływ korekcji PID na przykładzie charakterystyk częstotliwościowych.

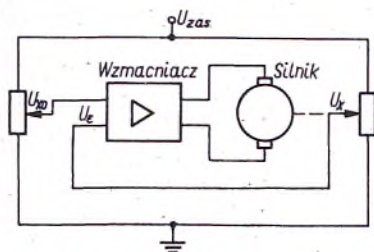
• 3.31• Czy regulatory PID nadają się do wszystkich rodzajów układów regulacji?

Nie, regulatory te są typowe dla zastosowań w układach, które określiliśmy poprzednio jako układy regulacji przemysłowej, a więc nadają się one przede wszystkim do regulacji stałowartościowej obiektów (procesów) typu technologicznego, takich jak piece, kotły czy reaktory chemiczne, które charakteryzują się dużymi stałymi czasowymi (ewentualnie opóźnieniami) i które są narażone na zakłócenia typu dodatkowych niekontrolowanych oddziaływań na wejściu. Specyfika układów regulacji przemysłowej z regulatorami PID ma swoje odbicie również w powszechnie stosowanym sposobie rysowania schematu takich układów (rys. 3.31). Przez taki sposób rysowania akcentuje się, że istotnym sygnałem wejściowym układu zamkniętego jest zakłócenie, zaś sygnał wartości zadanej x_0 można traktować do pewnego stopnia jako parametr.

Istnieje wiele układów regulacji, których struktura jest całkowicie zindywidualizowana i których projektowanie trudno ująć w jakiegokolwiek schematy. Dużą grupę układów regulacji, w których można określić pewną ogólną metodykę analizy i projektowania, lecz w których nie stosuje się regulatorów PID, są serwomechanizmy.

• 3.32• Co to jest serwomechanizm?

Serwomechanizmem nazywa się układ regulacji nadążnej położenia obiektu o charakterze mechanicznym. Zadaniem serwomechanizmu jest zwykle „śledzenie” sygnału zadanego o postaci bliżej nie określonej, przy czym sygnał ten reprezentuje niewielkie moce, natomiast oddziaływanie na wielkość regulowaną obiektu wymaga dużych mocy (prędkości, przyspieszeń). Zakłócenia w serwomechanizmach z zasady

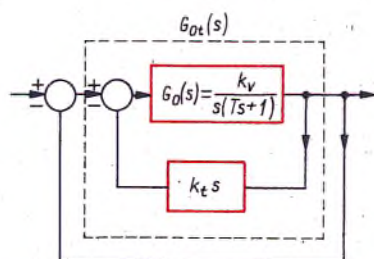


•3.32• Idea budowy układu serwomechanizmu

mają charakter zakłóceń pomiarowych sygnałów: zadanego i regulowanego — w konsekwencji zakłócenia pojawiają się w przebiegu uchybu regulacji, a więc na wejściu regulatora, a nie obiektu. Nieodłącznym elementem serwomechanizmu jest zwykle silnik wykonawczy (tzw. serwowmotor), od którego jakości zależy w sposób zasadniczy poprawna praca serwomechanizmu. Przykładem serwomechanizmu może być zarówno układ automatycznego śledzenia celu przez działo przeciwlotnicze, układ nastawiania położenia steru w dużym statku, układ regulacji położenia elektrody w stalowniczym piecu elektrycznym, czy wreszcie układ napędowy rejestratora napięcia (tzw. autokompensatora). Idea budowy serwomechanizmu jest przedstawiona na rys. 3.32.

•3.33• Jakie zadania są typowe dla serwomechanizmów?

Praca serwomechanizmów, a zwłaszcza ich projektowanie, musi być rozważona w kontekście zadań, do jakich są przeznaczone, ponieważ zadania te mogą być zupełnie różne. Podstawowym zadaniem serwomechanizmu jest zadanie naśladowania, czyli śledzenia zmiennego sygnału zadanego. Do rozwiązania tego zadania używa się modelu liniowego serwomechanizmu, w szczególności zaś metod częstotliwościowych, ponieważ określenie „nieznanych” sygnałów jest najprostsze za pomocą pojęć częstotliwościowych. Jeśli potrafimy oszacować widmo spodziewanych sygnałów zadanych oraz widmo zakłóceń pomiarowych, to serwomechanizm powinien zapewniać dobrą regulację w pasmie pokrywającym widmo sygnałów zadanych (a więc np. częstotliwościowy wskaźnik jakości regulacji $q(\omega)$ powinien być w tym pasmie dostatecznie mały), jednak poza tym pasmem roboczym serwomechanizm powinien wykazywać silne tłumienie (zakłóceń). Nieco inna jest sytuacja, gdy dokładnie jest określona klasa sygnałów zadanych, którą układ ma odtwarzać. Prowadzi to bowiem do wyraźnego sprecyzowania za pomocą pojęć astatyzmu dokładności w stanie ustalonym, ale także do określania żądanej postaci przebiegów czasowych. Klasyfikacyjnym zadaniem z tej grupy jest zadanie odtwarzania skokowych zmian wartości zadanej (lub reakcji na skokowe zakłócenie) czyli zadanie przestawiania. W zadaniu tym problem jest sformułowany następująco: jak można najlepiej sprowadzić do zera duży uchyb, jeśli taki się pojawi. Przez „najlepiej” niemal zawsze rozumie się tutaj — najszybciej. Zadanie to najwłaściwiej rozwiązuje się przyjmując model nieliniowy — z ograniczeniami mocy, prędkości lub przyspieszeń.



•3.34A• Schemat modelu serwomechanizmu ze sprzężeniem tachometrycznym

•3.34• Jak można przedstawić model najprostszego serwomechanizmu liniowego?

Układ otwarty najprostszego serwomechanizmu (rys. 3.32) składa się ze wzmacniacza i silnika. Transmitancja takiego układu ma uproszczoną postać

$$G_0(s) = \frac{k_v}{s(Ts+1)}$$

Obecność całkowania wynika z przyjęcia za wielkość wyjściową położenia, a nie prędkości kątowej silnika; współczynnik k_v określa wzmocnienie prędkościowe (patrz pyt. 2.53). Czynniki $(Ts+1)$ w mianowniku opisuje inercję (bezwładność) układu ruchowego, przejawiającą się tym, że po wystawieniu (np. skokowym) silnika jego prędkość kątowa nie ustali się od razu, lecz będzie się zmieniać stopniowo, ze stałą czasową T . Przy danej transmitancji układu otwartego wyrażenie na transmitancję układu zamkniętego jest następujące:

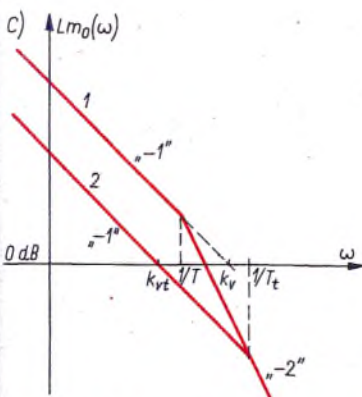
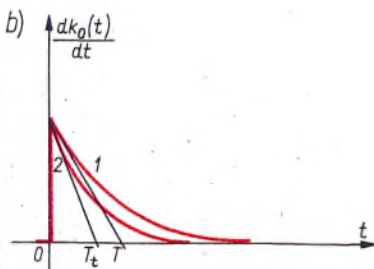
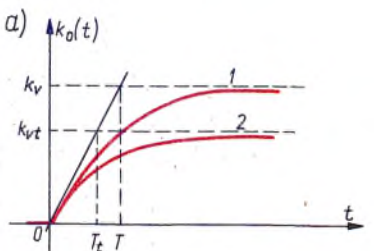
$$G(s) = \frac{G_0}{1+G_0} = \frac{1}{\frac{T}{k_v} s^2 + \frac{1}{k_v} s + 1} = \frac{1}{T_n^2 s^2 + 2\zeta T_n s + 1}$$

Transmitancja ta ma więc znaną postać transmitancji członu oscylacyjnego, przy czym $T_n = \sqrt{\frac{T}{k_v}}$ oraz $\zeta = \frac{1}{2\sqrt{Tk_v}}$

Korzystne właściwości takiego układu uzyskuje się przy małych wartościach T_n i niezbyt małych wartościach ζ . Parametr T jest dany (jako parametr obiektu — silnika), natomiast parametr k_v można zmieniać za pośrednictwem wzmacniacza. Zwiększanie k_v powoduje co prawda zmniejszenie okresu drgań (przez zmniejszenie T_n), ale jednocześnie zmniejszenie współczynnika tłumienia. Bardzo skutecznym sposobem poprawy właściwości serwomechanizmu jest zastosowanie korekcji przez sprzężenie zwrotne (ujemne) **tachometryczne**, czyli sprzężenie zależne od prędkości. Prędkość jest pochodną położenia, ale rzadko w praktyce jest potrzebny specjalny człon różniczkujący dla uzyskania tego sygnału, ponieważ pomiar prędkości kątowej jest dość prosty przy użyciu tzw. prądnicy tachometrycznej sprzężonej mechanicznie z silnikiem i wytwarzającej, z bardzo dobrym przybliżeniem, napięcie proporcjonalne do prędkości kątowej. Schemat modelu najprostszego serwomechanizmu ze sprzężeniem tachometrycznym przedstawiono na rys. 3.34A, przy czym łatwo zauważyć, że „pętla wewnętrzna”, uwzględniająca tylko sprzężenie tachometryczne, ma transmitancję $G_{or}(s)$ o identycznej postaci jako pierwotny układ otwarty, lecz tylko o innych parametrach liczbowych, mianowicie

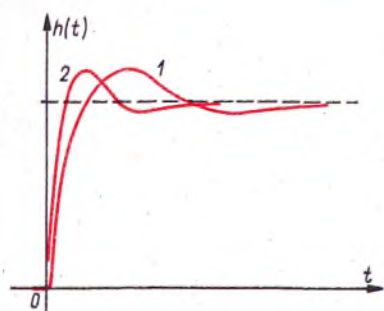
$$G_{or}(s) = \frac{G_0(s)}{1+k_t s G_0(s)} = \frac{k_{vt}}{s(T_t s+1)}$$

przy czym $k_{vt} = k_v/(1+k_v k_t)$, $T_t = T/(1+k_v k_t)$. Wpływ sprzężenia tachometrycznego o współczynniku k_t na kształt charakterystyk czasowych i częstotliwościowych układu otwartego jest widoczny na rys. 3.34B. Należy zwrócić uwagę na zmniejszenie k_{vt} i T_t w identycznym stosunku względem k_v i T oraz na niezmienność wartości pochodnej w chwili $t=0$. Ponieważ zmniejszeniu w jednakowym stopniu ule-



• 3.34B. Wpływ sprzężenia tachometrycznego na układ otwarty serwomechanizmu: a) charakterystyka impulsowa; b) pochodna charakterystyki impulsowej; c) charakterystyka częstotliwościowa amplitudowa

1 — układ bez sprzężenia tachometrycznego;
2 — ze sprzężeniem tachometrycznym



•3.34C• Charakterystyki skokowe serwo-mechanizmów dla identycznych współczynników tłumienia i różnych okresów drgań własnych

1 — okres drgań własnych dwa razy większy niż w przebiegu 2

gły dwa parametry „przeciwstawne” k_v i T , jest to korzystne dla parametrów pełnego układu zamkniętego

$$T_{nt} = \sqrt{\frac{T}{k_v}} = T_n;$$

$$\zeta_t = \frac{1}{2\sqrt{T_i k_{vt}}} = (1 + k_v k_t) \zeta$$

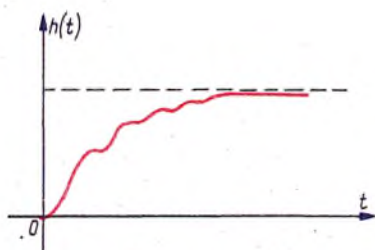
A więc zastosowanie korekcji przez sprzężenie tachometryczne umożliwia, przy stałych wartościach k_v i T , uzyskanie takiego samego okresu drgań własnych przy zwiększonym współczynniku tłumienia. Z drugiej strony, utrzymanie identycznego współczynnika tłumienia w układzie ze sprzężeniem tachometrycznym wyraźnie zwiększa współczynnik k_v , a więc przyspiesza przebiegi — zmniejsza T_n . Wynika to z zależności: jeśli przyjąć $\zeta_t = \zeta$, czyli

$$\frac{1 + k_v k_t}{2\sqrt{T k_v}} = \frac{1}{2\sqrt{T k_{v0}}}$$

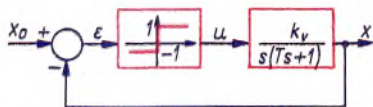
przy czym k_{v0} oznacza współczynnik wzmocnienia w układzie bez sprzężenia tachometrycznego, to przy $(1 + k_v k_t) = a > 1$ można przyjąć $k_v = a^2 k_{v0}$. Odpowiada to jednak a -krotnemu zmniejszeniu T_n . Na przykład jeśli pierwotny układ serwo-mechanizmu ma parametry T_{n0} i ζ_0 , przy danych T i k_{v0} , to można łatwo przyjąć $k_v = 4k_{v0}$, $k_t = 0,25$ — tak, że $1 + k_v k_t = 2$ i w układzie z takim sprzężeniem tachometrycznym uzyska się $T_n = \frac{1}{2} T_{n0}$ przy identycznym tłumieniu ζ_0 ; zilustrowano to na rys. 3.34C.

•3.35• Czy sprzężenie tachometryczne jest skuteczne tylko w prostym układzie liniowym drugiego rzędu?

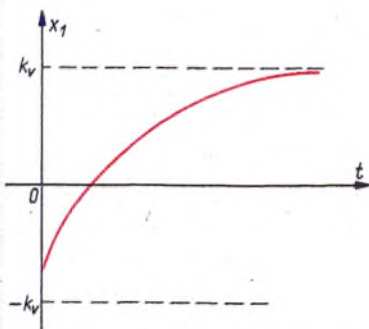
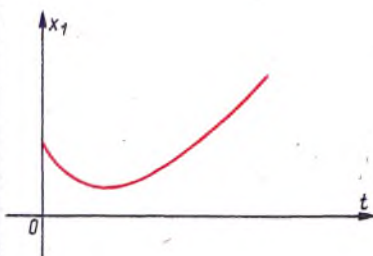
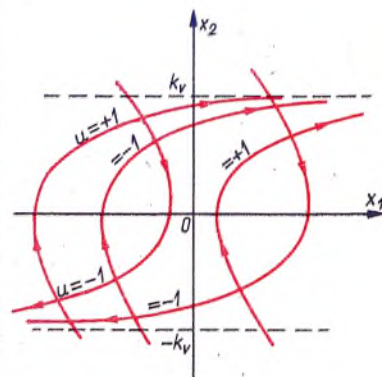
Nie, idea korekcji dynamicznej za pomocą sprzężenia tachometrycznego ma szerokie zastosowanie. Przykład z poprzedniego pytania wyjaśnia tylko sens tej korekcji, przy czym stosuje się ona do znacznie szerszej klasy układów — praktycznie nie buduje się serwo-mechanizmów bez sprzężenia tachometrycznego. Omówiony poprzednio najprostszy model liniowy drugiego rzędu jest „pierwszym przybliżeniem” opisu wszystkich układów rzeczywistych, wykazujących większą różnorodność zjawisk. Wystarczy np. aby rząd układu był wyższy niż 2 (sytuacja taka ma miejsce już w przypadku, gdy sprzężenie tachometryczne nie jest „idealne” — zamiast $k_t s$ występuje $k_t \frac{s}{1 + \tau s}$) — a łatwo uzyskać przebiegi o postaci jak na rys. 3.35, niemożliwe do uzyskania w modelu drugiego rzędu. Przebiegi takie otrzymuje się przy dużych wartościach k_t — sprzężenie tachometryczne wywołuje silne tłumienie „średnie”, ale jednocześnie przyczynia się do powstawania składowej oscylacyjnej.



•3.35• Charakterystyka skokowa serwo-mechanizmu możliwa do uzyskania tylko w układzie rzędu wyższego niż 2



·3.36A· Schemat blokowy modelu elementarnego serwomechanizmu przekąźnikowego



·3.36B· Trajektorie fazowe (a) i przebiegi czasowe (b) silnika z ograniczeniem prędkości przy sterowaniu $u = \pm 1$

·3.36· Jak można przedstawić elementarny model serwomechanizmu przekąźnikowego?

Idea serwomechanizmu przekąźnikowego stanowi dość naturalne rozwinięcie idei prostego serwomechanizmu liniowego (pyt. 3.34), jeśli rozważymy go pod względem zadania przestawiania. Zadanie takie sugeruje potrzebę odtwarzania przez serwomechanizm wymuszeń skokowych w możliwie krótkim czasie, co pociągałoby za sobą konieczność zwiększania wzmocnienia prędkościowego k_v . Z drugiej strony, rzeczywiste układy techniczne nie dopuszczają możliwości zwiększania żadnego sygnału nieograniczenie. Element przekąźnikowy stanowi rozwiązanie sytuacji, ponieważ skokowość przekąźnikowej charakterystyki statycznej można interpretować jako nieskończoność wzmocnienia, zaś ograniczoność tej charakterystyki — jako ograniczoność sygnałów fizycznych. Elementarny serwomechanizm przekąźnikowy można przedstawić tak jak na rys. 3.36A. Charakterystyka elementu przekąźnikowego jest unormowana do poziomów ± 1 , bowiem faktyczny poziom ograniczenia można określić za pomocą k_v . Ograniczenie to oczywiście należy interpretować jako ograniczenie prędkości — przy sterowaniu silnika o transmitancji $k_v/s(Ts+1)$ za pośrednictwem elementu przekąźnikowego nie można osiągnąć prędkości przekraczających zakres $\pm k_v$. Układ o takiej postaci nadaje się bardzo dobrze do analizy metodą płaszczyzny fazowej, można przy tym dla wygody przyjąć współrzędne fazowe

$$x_1 = \varepsilon = x_0 - x$$

$$x_2 = \dot{\varepsilon} = -\dot{x} \quad \text{przy } x_0 = \text{const}$$

Przyjęcie takich współrzędnych oznacza ustalenie linii połączeń przekąźnika przy $x_1 = 0$ — tzn. pokrywającej się z osią x_2 . Trajektorie w prawej i lewej półpłaszczyźnie mają kształt podobny, różnią się tylko znakiem sygnału sterującego u . Równania stanu mają postać

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = \frac{1}{T} (-x_2 - k_v u)$$

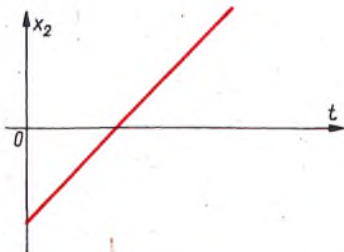
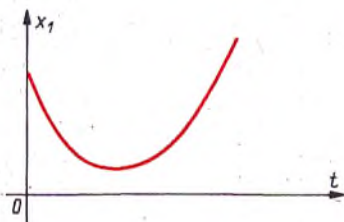
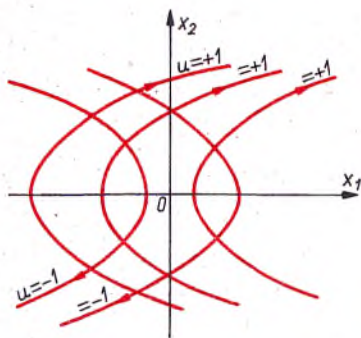
przy czym $u = \text{sign } x_1$, tzn. $u = +1$ dla $x_1 > 0$ lub -1 dla $x_1 < 0$. Równanie pojedynczej trajektorii fazowej można więc zapisać

$$\frac{dx_2}{dx_1} = -\frac{x_2 \pm k_v}{Tx_2}$$

Po rozdzieleniu zmiennych i scałkowaniu otrzymuje się

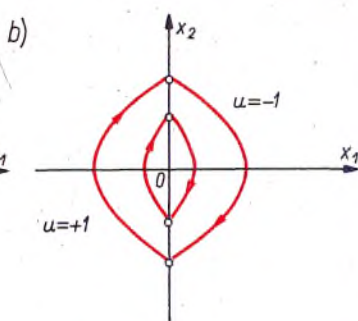
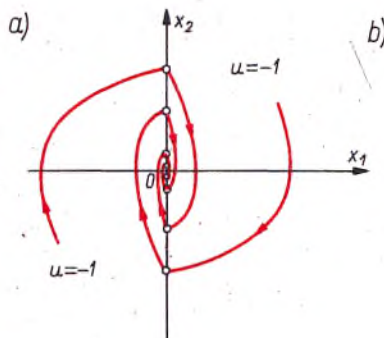
$$x_1 - x_1^0 = -T(x_2 - x_2^0) \pm Tk_v \ln |(x_2 \pm k_v)/(x_2^0 \pm k_v)|$$

przy czym x_1^0 i x_2^0 oznaczają współrzędne warunków początkowych, zaś znaki „ $\pm$ ” odpowiadają znakowi sygnału



•3.36C• Trajektorie fazowe (a) i przebiegi czasowe (b) silnika z ograniczeniem przyspieszenia przy sterowaniu $u = \pm 1$

•3.36D• Portrety fazowe elementarnego serwomechanizmu przekątnikowego: a) silnik z ograniczeniem prędkości; b) z ograniczeniem przyspieszenia



u . Trajektorie tego typu i przebiegi czasowe $x_1(t)$ i $x_2(t)$ pokazano na rys. 3.36B. Przy $k_v \rightarrow \infty$ i $T \rightarrow \infty$, ale $k_v/T = \text{const} = k_a$, co odpowiada ograniczeniu przyspieszenia występującemu np. przy sterowaniu silnika prądu stałego prądem twornika, otrzymamy w granicznym przypadku równania wyidealizowane

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -k_a u; \quad u = \text{sign } x_1$$

Łatwo sprawdzić, że trajektorie są wtedy parabolami (rys. 3.36C)

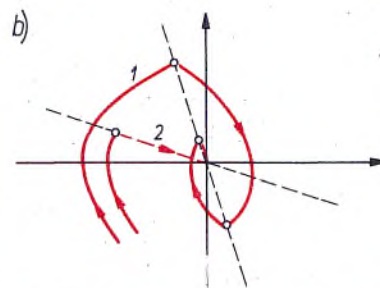
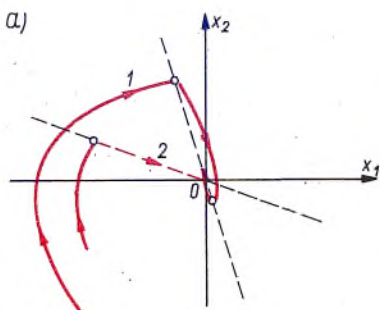
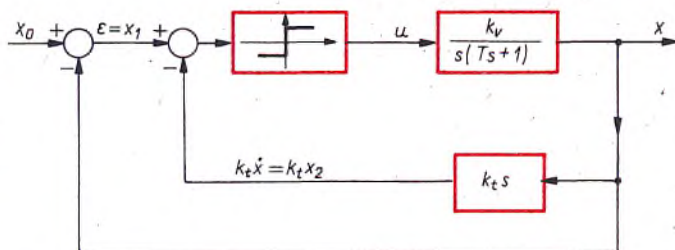
$$x_1 - x_1^0 = \pm \frac{1}{2k_a} [x_2^2 - (x_2^0)^2]$$

Na rysunku 3.36D pokazano portrety fazowe układu zamkniętego takiego serwomechanizmu; portrety powstają przez łączenie na linii przełączeń $x_1 = 0$ trajektorii z dwu rodzin przy $u = \pm 1$. Jak widać, portrety fazowe wykazują wielką niedoskonałość takich układów. Nic w tym dziwnego, ponieważ zastosowane tu sterowanie przekątnikowe jest uzależnione tylko od zmiennej x_1 , a przełączenia przy zmianie znaku x_1 następują zbyt późno na to, aby uzyskać trajektorie szybciej zmierzające do stanu równowagi $(0, 0)$. W przypadku (a) układ jest co prawda stabilny asymptotycznie, ale w przypadku (b) punkt równowagi jest typu środka — układ jest na granicy stabilności.

•3.37• Jak można najprościej skorygować działanie serwomechanizmu przekątnikowego?

Podobnie jak w przypadku liniowym — radykalną metodą korekcji jest sprzężenie tachometryczne wprowadzone tym razem na wejście elementu przekątnikowego (rys. 3.37A). Skoro teraz na wejściu przekątnika występuje sygnał $x_1 + k_t x_2$, to wobec tego sterowanie u jest określone zależnością $u = \text{sign}(x_1 + k_t x_2)$, a stąd linię przełączeń opisuje zależność $x_2 = -x_1/k_t$ (a nie $x_1 = 0$ jak poprzednio). Rysunek 3.37B przedstawia portrety fazowe układów ze sprzężeniem tachometrycznym. Przy k_t małym (a więc przy sprzę-

•3.37A• Schemat blokowy modelu serwomechanizmu przekąźnikowego ze sprzężeniem tachometrycznym



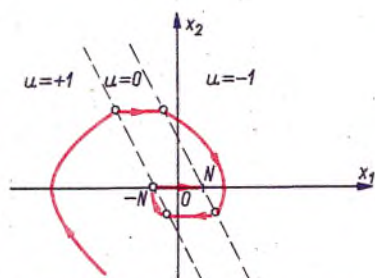
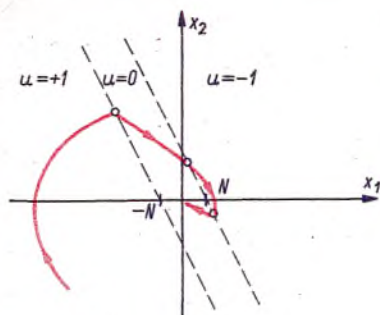
•3.37B• Portrety fazowe serwomechanizmów przekąźnikowych ze sprzężeniem tachometrycznym: a) układ z ograniczeniem prędkości; b) z ograniczeniem przyspieszenia

1 — sprzężenie tachometryczne słabe; 2 — sprzężenie silne

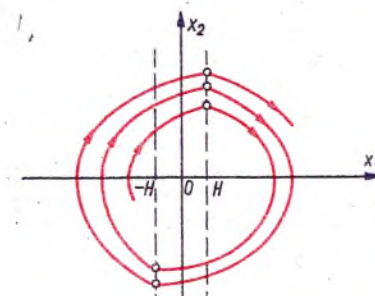
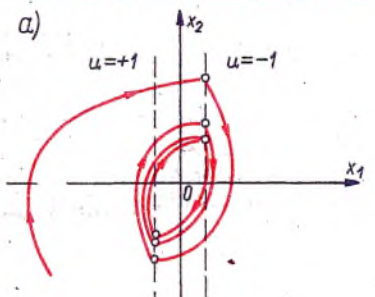
zeniu tachometrycznym „słabym” i linii przełączeń niewiele odchylonej od osi x_2 uzyskuje się od razu widoczne polepszenie charakteru trajektorii, które wykazują mniejszą skłonność do oscylacji (zwłaszcza w przypadku (b)) — jest to analogia do zwiększenia tłumienia w przypadku liniowym. Natomiast przy k_t bardzo dużym (a więc przy sprzężeniu tachometrycznym „silnym” i linii przełączeń mocno nachylonej) otrzymuje się odpowiednik przetłumienia w przypadku liniowym. Tutaj jednak należy to zinterpretować nieco inaczej: przy linii przełączeń mocno nachylonej trajektorie nie mogą być przełączane na jedną lub drugą stronę linii przełączeń, lecz następuje niejako „wymuszanie” ruchu po samej linii przełączeń. Zjawisko tego typu nazywa się **ruchem ślizgowym**. Można też ująć to inaczej: przy bardzo dużych wartościach k_t wewnętrzną pętlę sprzężenia tachometrycznego można w całości zastąpić elementem zlinearyzowanym bezinercyjnym o wzmacnieniu $1/k_t$, ponieważ zawsze przy bardzo dużym wzmacnieniu w pętli sprzężenia zwrotnego sygnał za węzłem sumującym jest bliski zeru (w tym przypadku $x_1 + k_t x_2 \approx 0$, stąd $x_2 = -x_1/k_t$). Sens takiej linearyzacji jest identyczny jak w przypadku omówionym w pyt. 2.33, gdzie odwołano się do faktycznej „dynamiczności” zagadnienia — tutaj jednak obecność inercji $k_v/(T_s + 1)$ umożliwia jednoznaczną interpretację linearyzacji charakterystyki w rozumieniu średnim w czasie. Pośrednie wartości sygnału u (pomiędzy $+1$ i -1), umożliwiające utrzymanie trajektorii o innym kształcie niż przy $u = \pm 1$, należy rozumieć jako wartości średnie przebiegu o postaci przemiennie przełączanych wartości $+1$ i -1 , przy czym przełączenia te odbywają się teoretycznie nieskończenie często. Zjawisko ruchu ślizgowego nie jest korzystne, ponieważ wymuszanie trajektorii o kształcie linii prostej odpowiada wydłużaniu (do nieskończoności) czasu dojścia do położenia równowagi, a poza tym „nieskończenie częste” przełączanie zazwyczaj nie jest zbyt korzystne dla elementu przekąźnikowego, zwłaszcza jeśli jest to rzeczywisty element łączeniowy.

•3.38• Jaki wpływ na działanie serwomechanizmu ma „realność” elementu przekąźnikowego?

W poprzednich dwu pytaniach założono, że element przekąźnikowy jest idealny i opisuje się zależnością typu funkcji



•3.38A• Portrety fazowe serwomechanizmów przekąźnikowych przy występowaniu strefy nieczułości przekąźnika: a) układ z ograniczeniem prędkości; b) z ograniczeniem przyspieszenia



•3.38B• Portrety fazowe serwomechanizmów przekąźnikowych przy występowaniu histerezy przekąźnika: a) układ z ograniczeniem prędkości; b) z ograniczeniem przyspieszenia

signum. W rzeczywistości element przekąźnikowy charakteryzuje się strefą nieczułości albo histerezą (do pewnego stopnia zjawisku histerezy jest równoważne opóźnienie czasowe działania przekąźnika) — najczęściej zaś trzeba się liczyć z kilkoma zjawiskami jednocześnie.

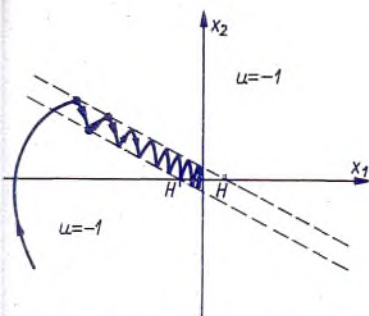
Strefa nieczułości jest łatwa do uwzględnienia, trzeba tylko wziąć pod uwagę jeszcze jeden typ trajektorii fazowych — przy $u = 0$. Trajektorie takie mają kształt liniowy: $x_2 - x_2^0 =$

$\frac{1}{T} (x_1 - x_1^0)$ w przypadku ograniczenia prędkości i $x_2 = x_2^0$ — w przypadku ograniczenia przyspieszenia. Odpowiednie portrety fazowe serwomechanizmu ze sprzężeniem tachometrycznym i ze strefą nieczułości przedstawiono na rys. 3.38A. Przy braku sprzężenia tachometrycznego sytuacja byłaby podobna. Należy zwrócić uwagę na możliwość otrzymania uchybu w stanie ustalonym w granicach strefy nieczułości.

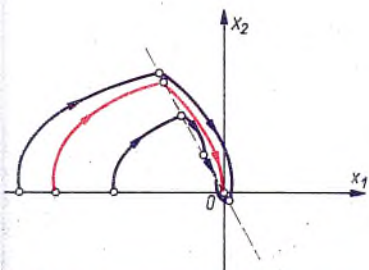
Zjawisko histerezy (a także opóźnienia) ma konsekwencje bardziej istotne. Przy braku sprzężenia tachometrycznego powoduje ono destabilizację układu przez „rozsunięcie” linii przełączeń w niekorzystnym kierunku (rys. 3.38B), przy umiarkowanym sprzężeniu tachometrycznym „osłabia” w pewnym stopniu jego działanie korekcyjne. Natomiast przy bardzo silnym sprzężeniu tachometrycznym efekt histerezy podkreśla w specyficzny sposób zjawisko ruchu ślizgowego. Na rysunku 3.38C pokazano przykłady trajektorii dla takiego przypadku. Linia przełączeń uległa rozszczepieniu na dwie równoległe linie, przy czym ruch pomiędzy tymi liniami jest niejednoznaczny — można to sobie wyobrazić jako dwie warstwy połączone „na zakładkę”, przy czym o przynależności trajektorii do danej warstwy decyduje poprzedni przebieg trajektorii. Tak więc zjawisko histerezy wprowadza jak gdyby trzecią zmienną stanu (tylko dwuwartościową) — mianowicie zmianę stanu przekąźnika, która jest określona nie tylko przez wielkość na wejściu tego przekąźnika, ale i przez historię ruchu. W konsekwencji otrzymujemy ruch ślizgowy w postaci wyraźnych przełączeń o skończonym czasie, przy czym nadal obowiązuje zasada „średnich w czasie” wartości u , związanych ze „średnią” trajektorią wzdłuż linii przełączeń. Sytuacja ta stanowi pewną analogię do przypadku układu liniowego i przebiegów z rys. 3.35. Jest to oczywiście przypadek niekorzystny.

•3.39• Jak najlepiej można zrealizować zadanie przestawiania?

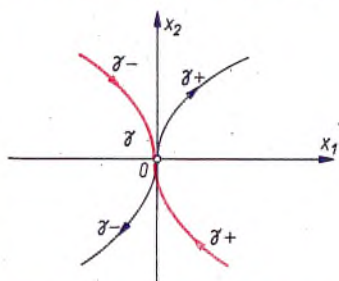
Na podstawie pytań 3.36 ... 3.38 można podjąć próbę intuicyjnego uzasadnienia idei sterowania czasoptymalnego, bo takie właśnie sterowanie jest najlepsze dla realizacji zadania przestawiania. Korzystając z przykładu prostego serwomechanizmu drugiego rzędu można zauważyć, że „najlepsze” trajektorie pod względem czasu przestawiania jest w stanie zapewnić układ przekąźnikowy, ponieważ pracuje



•3.38C• Portret fazowy serwomechanizmu przekąźnikowego ze sprzężeniem tachometrycznym (silnym) przy występowaniu histerezy przekąźnika — ruch ślizgowy

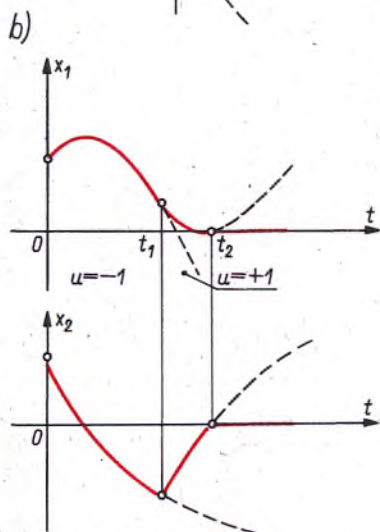
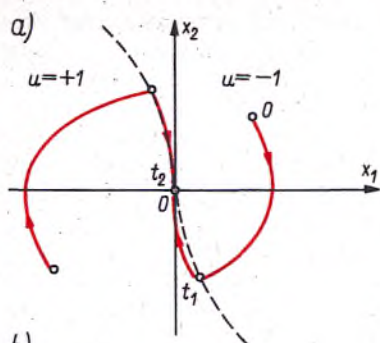


•3.39A• Kilka trajektorii fazowych serwomechanizmu przekąźnikowego ze sprzężeniem tachometrycznym dla różnych warunków początkowych $x_1(0)$



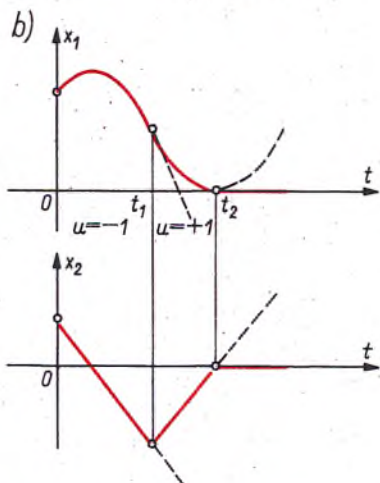
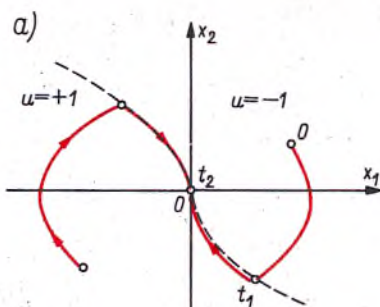
•3.39B• Sposób konstrukcji linii przełączeń serwomechanizmu czasooptymalnego z dwóch odcinków trajektorii przechodzących przez punkt $(0,0)$

on przy największych prędkościach lub przyspieszeniach dopuszczalnych w danym serwo motorze. Spełnienie zadania przedstawiania wymaga jednak, aby trajektorie osiągały zadany cel — punkt $(0,0)$ na płaszczyźnie fazowej — przy dowolnych warunkach początkowych. Serwomechanizm bez sprzężenia tachometrycznego nie może być oczywiście brany pod uwagę. Układ ze sprzężeniem tachometrycznym (pyt. 3.37) zapewnia osiągnięcie celu, jednak z rys. 3.39A wynika, że na ogół odbywa się to w sposób nieoptymalny. Przy wartościach $x_1(0)$ (uchybu w chwili początkowej) bardzo oddległych od punktu $(0,0)$ trajektoria ma kilka punktów przełączeń, co powoduje wydłużanie czasu osiągnięcia celu. Z kolei przy małych wartościach $x_1(0)$ — duży odcinek jest przybierany na zasadzie ruchu ślizgowego, co również (patrz pyt. 3.37) powoduje wydłużenie czasu. Natomiast trajektoria wyróżniona na rysunku grubszą linią jest czasooptymalna: pierwszy jej odcinek odpowiada maksymalnemu „rozpędzaniu” układu, zaś drugi — maksymalnemu „hamowaniu”, przy czym ten ostatni odcinek trafia dokładnie w punkt $(0,0)$, w którym to punkcie obydwie współrzędne stanu, równe zero, powodują, że i stan elementu przekąźnikowego (rzeczywistego) jest równy 0, zatem ruch zostaje zatrzymany. Nasuwa się jednak od razu pytanie, jak sprawić, aby wszystkie trajektorie, a nie tylko jedna wyróżniona, podlegały podobnym prawom. Uważna analiza portretu fazowego serwomechanizmu przekąźnikowego pozwala wysnuć wniosek, który można uzasadnić formalnie (czego tutaj nie będziemy robić). Najważniejszym spostrzeżeniem jest to, że przez punkt $(0,0)$ przechodzą tylko dwie trajektorie γ^- i γ^+ odpowiadające „maksymalnemu hamowaniu” — rys. 3.39B. Jeśli działanie układu ma zapewnić minimum czasu, to dojście do punktu $(0,0)$ po jakiegokolwiek innej trajektorii w końcowym etapie stanowiłoby błędny sposób sterowania. Na to jednak, aby zawsze końcowy etap działania był zgodny z trajektorią „hamowania”, potrzeba i wystarcza, aby linia γ , złożona z dwóch odcinków γ^- i γ^+ pokrywających się z trajektoriami hamowania przechodzącymi przez punkt $(0,0)$, była linią przełączeń serwomechanizmu. Jak już zauważyliśmy w poprzednich pytaniach, linia przełączeń jest w istocie odwzorowaniem charakterystyki sprzężenia tachometrycznego. Tak więc serwomechanizm czasooptymalny, realizujący w najkrótszym czasie zadanie przedstawiania, jest serwomechanizmem przekąźnikowym z nieliniowym sprzężeniem tachometrycznym. Nieliniowość charakterystyki sprzężenia tachometrycznego jest podyktowana rodzajem trajektorii obiektu — w rozważanych przez nas przypadkach, byłyby to trajektorie z rys. 3.36B lub 3.36C (z uwzględnieniem złożenia dwóch odcinków γ^- i γ^+ , ponieważ linia przełączeń musi zachowywać symetrię względem początku układu współrzędnych). Na rysunkach 3.39C i 3.39D przedstawiono portrety fazowe serwomechanizmu czasooptymalnego dla dwóch rozważanych wersji modelu silnika



• 3.39C • Portret fazowy (a) i przebiegi czasowe (b) w serwo mechanizmie czasooptymalnym z ograniczeniem prędkości

• 3.39D • Portret fazowy (a) i przebiegi czasowe (b) w serwo mechanizmie czasooptymalnym z ograniczeniem przyspieszenia



wraz z przebiegami czasowymi interesujących zmiennych. Jeśli nawet rzeczywisty układ serwo mechanizmu z różnych względów (nieidealność charakterystyk elementów, błędny dobór typu lub parametrów modelu obiektu) nie byłby ściśle optymalny, to przedstawione pojęcia stanowią wzorzec wskazujący kierunek postępowania przy syntezie układu realizującego czasooptymalnie zadanie przestawiania.

·4.1· Jakie urządzenia są potrzebne do budowy układów regulacji automatycznej?

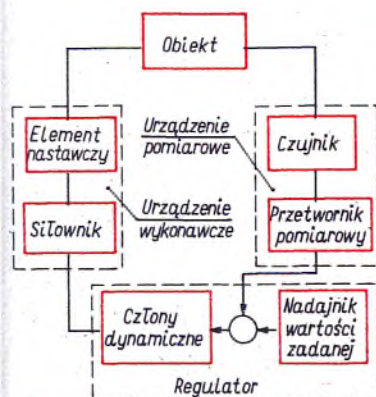
Zestaw urządzeń potrzebnych do zbudowania układu regulacji automatycznej można podać dla pewnej określonej klasy układów. Na rysunku 4.1. przedstawiono schemat blokowy jednopętlowego układu regulacji stałowartościowej. Ten typ regulacji jest spotykany najczęściej. Bardziej skomplikowane układy zawierają szereg układów jednopętlowych, dla których wartości zadane mogą być zmieniane przez inne układy.

Na schemacie wyodrębniono urządzenia pomiarowe, wykonawcze oraz regulator. Urządzenia pomiarowe są to czujniki i przetworniki pomiarowe. Do urządzeń wykonawczych zaliczamy elementy nastawczy, czyli wszelkiego rodzaju zawory i zasuwy oraz elementy wykonawcze czyli siłowniki. Regulatory są najważniejszymi urządzeniami spośród tzw. urządzeń części centralnej. Do tej grupy należą także przystawki matematyczne wykonujące dodawanie, mnożenie, dzielenie, pierwiastkowanie oraz rejestratory i przetworniki zmieniające fizyczną postać sygnału, np. przetworniki elektro-pneumatyczne, pneumatyczno-elektryczne i inne.

·4.2· Według jakich zasad jest prowadzona standaryzacja sprzętu automatyki?

Automatyzacji podlega wiele bardzo różnorodnych procesów technologicznych. Powoduje to zwiększenie zapotrzebowania na pomiary wielu różnych wielkości fizycznych przy różnych zakresach ich zmienności. Z drugiej strony, prawo (algorytm), według którego powinna przebiegać regulacja, nie zależy od fizycznej realizacji obiektu, a jedynie od jego transmitancji. Wynika stąd, że można tworzyć układy regulacji dla wielu różnych procesów używając takiej samej aparatury, pod warunkiem, że urządzenia mające bezpośrednią styczność z obiektem (tzn. czujniki pomiarowe i ew. elementy wykonawcze) będą dobrane zgodnie z wymaganiami automatyzowanego procesu.

Standaryzacja aparatury regulacyjnej ma na celu umożliwie-



·4.1· Schemat blokowy jednopętlowego układu regulacji

nie tworzenia różnorodnych układów regulacji przy użyciu pewnej liczby urządzeń standardowych (uniwersalnych) i pojedynczych urządzeń specjalizowanych mających styczność z obiektem. Powstały tzw. systemy blokowe aparatury regulacyjnej. Urządzenia wchodzące w skład tych systemów można zestawiać, zależnie od potrzeb, w różnych konfiguracjach, tworząc układy o różnym przeznaczeniu.

4.3. Jakie parametry urządzeń regulacji podlegają standaryzacji w ramach systemów blokowych?

Wymagania stawiane urządzeniom wchodzącym w skład systemów blokowych mają na celu przede wszystkim umożliwienie łączenia ze sobą poszczególnych bloków. Stwarza to konieczność ujednolicenia sygnałów, tak aby sygnały wyjściowe i wejściowe wszystkich bloków były identyczne co do postaci, zakresów zmienności, mocy i innych parametrów. Cechą podstawową systemów blokowych jest więc posługiwanie się standardowym sygnałem roboczym. Ponadto standaryzacji wewnątrz systemów blokowych podlegają wymiary, sposoby montażu, sposoby łączenia, wartości ciśnień i napięć zasilania oraz ewentualnie inne parametry techniczne.

Dla aparatury produkowanej w Polsce przyjęto jako sygnały standardowe: elektryczny 0 ... 5 mA, 0 ... 20 mA (dla nowo opracowanej aparatury), 4 ... 20 mA; pneumatyczny: 19 613 ... 98 066,5 Pa (0,2 ... 1 kG/cm²). Podobne standardy zostały przyjęte przez firmy zagraniczne. Dla aparatury pneumatycznej najczęściej są stosowane sygnały 0,2 ... 1 kG/cm², a dla aparatury elektrycznej — sygnały 4 ... 20 mA. Posługiwanie się takim sygnałem standardowym umożliwia współpracę urządzeń polskich z zagranicznymi.

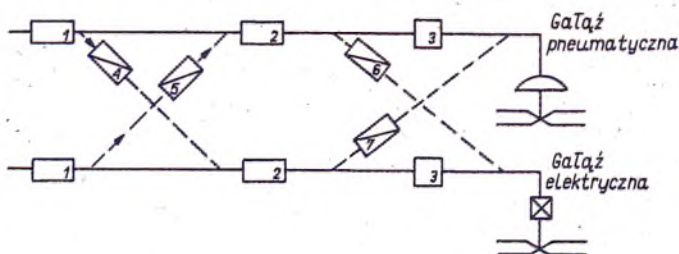
Standaryzacja prowadzi do zmniejszenia liczby typów potrzebnych urządzeń regulacyjnych i, w związku z tym, do wydłużenia ich serii produkcyjnych. Wydłużenie serii powoduje obniżenie ceny pojedynczego aparatu i całkowicie wynagradza ograniczenie swobody w doborze parametrów urządzeń, związane czasem z koniecznością pewnej rozbudowy aparaturowej układu regulacji.

4.4. Czy urządzenia należące do różnych systemów blokowych mogą ze sobą współpracować?

Tak. Większość systemów blokowych zawiera przetworniki zamieniające jeden sygnał standardowy na inny. Na rysunku 4.4 przedstawiono schematycznie sposób tworzenia układów złożonych z bloków jednej grupy (np. elektrycznej) oraz układów mieszanych. Połączenia między blokami jednego systemu blokowego narysowano liniami ciągłymi, a między blokami różnych systemów liniami przerywanymi. Ze schematu widać, że przetworniki międzysystemowe stanowią jak gdyby „klucze” otwierające drogę między systemami pracującymi przy różnych sygnałach standardowych. Takie

4.4. Możliwe warianty współpracy gałęzi pneumatycznej i elektrycznej blokowego systemu aparatury automatyki

1 — urządzenia pomiarowe; 2 — regulatory; 3 — wzmacniacze mocy; 4 — przetwornik pomiarowy pneumatyczno-elektryczny; 5 — przetwornik pomiarowy elektro-pneumatyczny; 6 — przetwornik pneumatyczno-elektryczny z wyjściem mocy; 7 — przetwornik elektro-pneumatyczny z wyjściem mocy

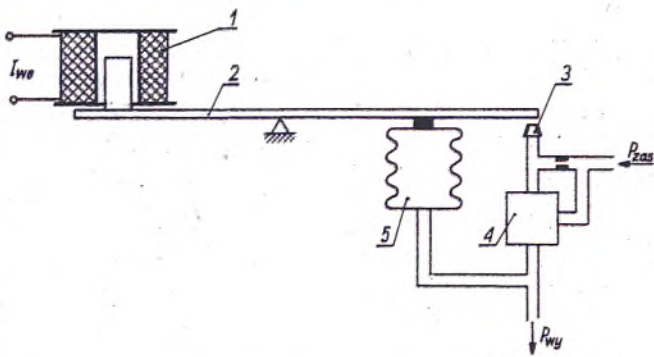


„klucze” umożliwiające dostęp do określonego systemu blokowego, tworzone są nie tylko w celu umożliwienia współpracy urządzeń wykorzystujących różne rodzaje energii pomocniczej. Przetworniki buduje się także w celu umożliwienia łączenia bloków, których sygnały standardowe mają tę samą postać, ale różne wartości. Tak np. przyjęty jako obowiązujący dla aparatury elektrycznej produkowanej w Polsce sygnał prądowy 0 ... 5 mA jest obecnie zastępowany sygnałem 4 ... 20 mA.

Przetworniki międzysystemowe elektro-pneumatyczne i pneumatyczno-elektryczne występują w większości systemów blokowych. Dzięki nim można tworzyć układy mieszane, w których np. urządzenia pomiarowe i przetwarzające sygnał są elektryczne, a urządzenia wykonawcze — pneumatyczne.

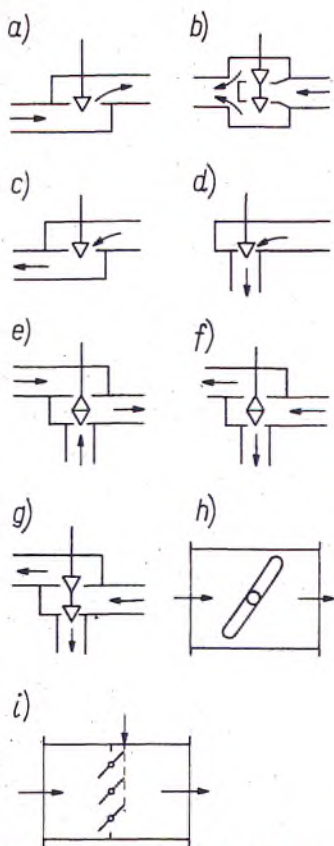
4.5. Jaki można podać przykład przetwornika międzysystemowego?

Budowę typowego międzysystemowego przetwornika elektro-pneumatycznego przedstawiono na rys. 4.5. Dla zwiększenia dokładności pracuje on jako układ autokompensacyjny, w układzie tzw. równoważni elektropneumatycznej. Prąd wejściowy przetwornika płynący przez cewkę 1 znajdującą się w polu magnesu trwałego powoduje zmianę położenia dźwigni 2, a więc zmianę ciśnienia w układzie dysza-przysłona, i stąd zmianę ciśnienia wyjściowego p_{wy} układu. Ciśnienie wyjściowe układu dysza-przysłona stanowi, po wzmacnieniu, sygnał wyjściowy przetwornika, a jednocześnie zasila mieszek 5 kompensujący siłę wytwarzaną przez



4.5. Przetwornik elektro-pneumatyczny

1 — cewka; 2 — dźwignia; 3 — zespół dysza-przysłona; 4 — wzmacniacz pneumatyczny; 5 — mieszek



4.6. Schematy konstrukcyjne zaworów i przepustnic: a) zawór jednogniazdowy; b) dwugniazdowy; c) jednogniazdowy wysokociśnieniowy; d) kątowy; e) trójdrożny mieszający; f, g) trójdrożne rozdzielające; h) przepustnica obrotowa; i) przepustnica żaluzjowa

elektromagnes. W stanie równowagi ciśnienie wyjściowe jest więc proporcjonalne do prądu wejściowego.

Zakres zmian prądu wejściowego jest zgodny ze standardowym zakresem przyjętym dla blokowego systemu aparatury elektrycznej, w skład którego wchodzi przetwornik, na wyjściu otrzymywany jest standardowy sygnał ciśnieniowy.

4.6. Jakiego rodzaju elementy nastawcze są stosowane w układach regulacji?

Elementy nastawcze umożliwiają oddziaływanie na przepływ energii lub materiału w kierunku do lub od obiektu. Zmieniając przepływ materiału lub energii można zmieniać przebiegi procesów zachodzących w obiekcie; elementy nastawcze umożliwiają więc sterowanie obiektów.

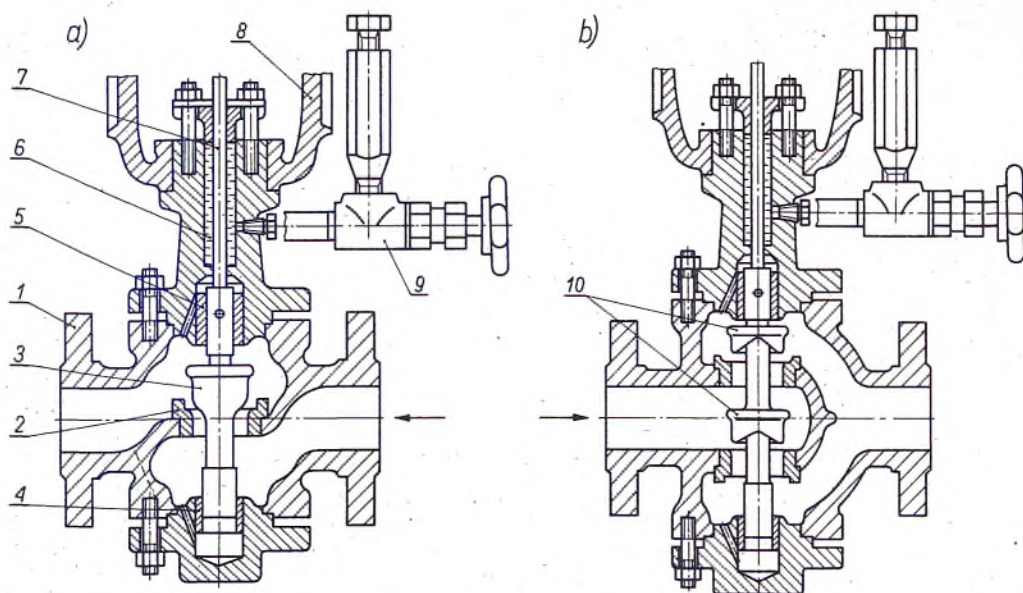
Najczęściej stosowanym rodzajem elementów nastawczych są zawory i zasuw. Służą one do sterowania natężenia przepływu cieczy, par lub gazów w przewodach rurowych. Schematy konstrukcyjne najczęściej używanych zaworów nastawczych przedstawiono na rys. 4.6. Zwykle zawory nastawcze dzielą się na zawory **jednogniazdowe** (rys. 4.6a) i **dwugniazdowe** (rys. 4.6b). Zawory jednogniazdowe wysokociśnieniowe, wykonywane jako **proste** (rys. 4.6c) lub **kątowe** (rys. 4.6d), różnią się od pozostałych budową oraz kierunkiem przepływu czynnika w zaworze. W zwykłych zaworach jednogniazdowych kierunek przepływu jest taki, że ciśnienie strumienia działa w kierunku otwierania zaworu, w zaworach wysokociśnieniowych jest odwrotnie. Zawory **trójdrożne**: mieszające i rozdzielające pokazano na rys. 4.6e, f, g.

W rurociągach o dużych przekrojach, do nastawiania natężenia przepływu cieczy par oraz gazów stosuje się kłapy dławiące zwane najczęściej **przepustnicami**. Na rys. 4.6h przedstawiono schemat konstrukcyjny przepustnicy obrotowej, a na rys. 4.6i przepustnicy żaluzjowej.

4.7. Jak jest zbudowany typowy zawór nastawczy?

Jako typowe przykłady zaworów nastawczych mogą posłużyć zawory produkcji zakładów MERA-POLNA w Przemyśle (rys. 4.7). W korpusie 1 zamocowane jest gniazdo zaworu 2, w które jest wsuwany grzybek zaworu 3. Powierzchnie gniazda i grzybka są dokładnie oszlifowane. Od dokładności obróbki tych dwu elementów oraz od zachowania ich współosiowości po zmontowaniu zależy **szczelność i charakterystyka otwarcia zaworu** (w skrócie charakterystyka zaworu). Grzybek zaworu może wykonywać ruch posuwisto-zwrotny.

W czasie swego ruchu grzybek jest utrzymywany w osi gniazda przez dwie prowadnice: dolną 4 i górną 5. Do poruszania grzybkiem służy, wyprowadzone na zewnątrz zaworu, wrzeciono 7. Wewnątrz zaworu może panować duże ciśnienie. Aby czynnik (gaz, ciecz lub para) znajdujący się wewnątrz zaworu nie wydostawał się na zewnątrz, wrzeciono przechodzi przez dławnicę z pakunkiem uszczelniającym 6.



4.7. Zawory regulacyjne (produkcji Zakładów „Polna”): a) zawór jednogniazdowy; b) zawór dwugniazdowy

1 — korpus; 2 — gniazdo; 3 — grzybek; 4 — prowadnica dolna; 5 — prowadnica górna; 6 — dławnica; 7 — wrzeciono; 8 — część korpusu siłownika; 9 — smarownica; 10 — podwójny grzybek zaworu dwugniazdowego

Zakłady POLNA wykonują pakunki uszczelniające z teflonu-azbestu, azbestu-grafitu lub teflonu. Dla zmniejszenia tarcia między pakunkiem uszczelniającym a wrzecionem zaworu, wrzeciono jest smarowane. Do smarowania służy smarownica śrubowa 9 z zaworem odcinającym. Dławnica służy nie tylko do uszczelniania zaworu, ale również zabezpiecza współpracujący z nim siłownik przed wpływem wysokiej temperatury czynnika.

W przedstawionych rozwiązaniach zaworów występują dwa rodzaje grzybków: konturowy i tłoczkowy. Pożądane charakterystyki otwarcia zaworów konturowych uzyskuje się dzięki odpowiednio dobranemu kształtowi grzybków, a pożądane charakterystyki zaworów tłoczkowych — dzięki odpowiedniemu kształtowi wycięt w ściankach tłoczków.

4.8. Co to są siłowniki i jakiego rodzaju siłowników używa się w układach regulacji automatycznej?

Siłownikami nazywa się elementy napędowe służące w układach regulacji automatycznej do ustawiania położenia zaworów i przepustnic. Aby spełnić to zadanie, elementy wyjściowe siłowników wykonują ruch prostoliniowy lub kątowy. Ważne są wartości następujących parametrów ruchu siłownika: **siła lub moment, czas przestawiania z jednego położenia skrajnego w drugie i wartość skoku**. Parametry różnego rodzaju siłowników zestawiono w tabl. 4.8. Stosując układy dźwigniowe można uzyskiwać większe wartości skoków przy mniejszych siłach i odwrotnie.

Z porównania danych zamieszczonych w tabl. 4.8. widać, że duże wartości skoków uzyskuje się w siłownikach tłokowych. One też umożliwiają uzyskiwanie największych wartości sił i momentów. Dla małych i średnich wartości sił

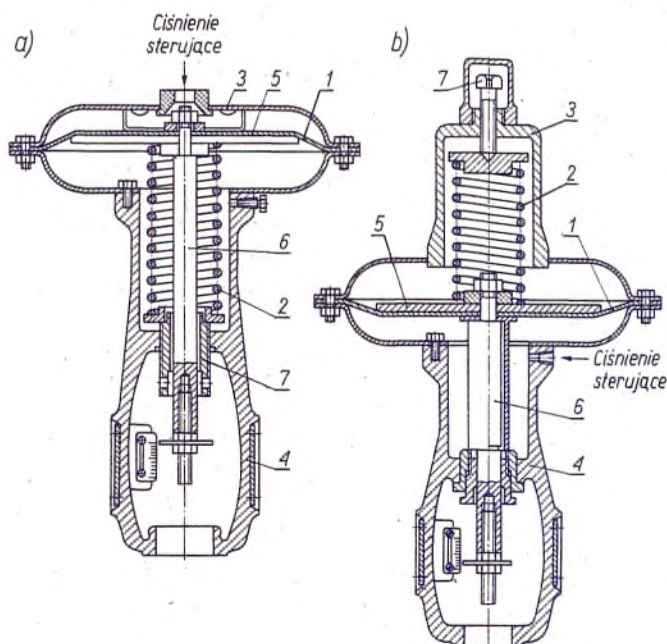
4.8. Parametry typowych siłowników

Rodzaj elementu	Siła [N]	Moment (dla siłowników kątowych) [N · m]	Skok [mm]	Skok [stopnie kątowe]	Czas przedstawiania [s]
Pneumatyczny membranowy	1000...10 000	500...5000	20...100	30	1...20
Pneumatyczny tłokowy	1000...30 000	500...15 000	60...500	90...110	5...50
Hydrauliczny tłokowy	2000...100 000	1000...50 000	100...1200	90...110	10...100
Elektryczny silnikowy (silnik o mocy 30...250 W)	1000...20 000	500...10 000	20...100	90...110	5...100

i skoku można stosować siłowniki dowolnego rodzaju. Wobec tego o doborze rodzaju siłownika decyduje najczęściej koszt, a więc prostota jego budowy. Tym właśnie tłumaczy się szybkie rozpowszechnienie siłowników pneumatycznych membranowych.

•4.9. Jak jest zbudowany i jaka jest zasada działania membranowego siłownika pneumatycznego?

Jako przykład budowy membranowego siłownika przedstawiono na rys. 4.9 typowy siłownik o działaniu prostym i siłownik o działaniu odwrotnym. W siłniku o działaniu prostym ciśnienie sterujące jest doprowadzane pod górną



•4.9. Siłowniki membranowe: a) prosty; b) odwrotny

1 — membrana; 2 — sprężyna; 3 — pokręta górną; 4 — obudowa dolna; 5 — talerz; 6 — trzpień; 7 — śruba regulacyjna

pokrywą siłownika. Działa ono na elastyczną (wykonaną np. z neoprenu) membranę 1, powodując jej uginanie. Membrana za pośrednictwem sztywnego talerza 5 naciska na sprężynę 2. Siła ściskanej sprężyny wzrasta proporcjonalnie do wartości ugięcia. Jeżeli więc sprężyna jest odpowiednio dobrana, to dla pewnej wartości ugięcia sprężyny następuje zrównoważenie siły wywieranej przez membranę przez siłę pochodzącą od ściskanej sprężyny. Dla każdej wartości ciśnienia równowaga następuje przy innej wartości ugięcia membrany. Ruch membrany jest przenoszony za pośrednictwem trzpienia 6 na zewnątrz obudowy siłownika. Wstępny naciąg sprężyny reguluje się przez wkręcanie w obudowę siłownika śruby 7.

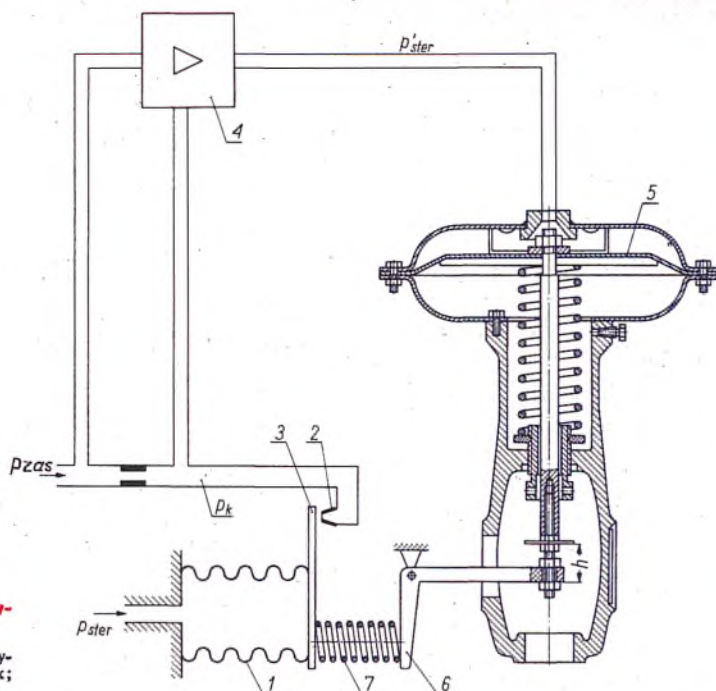
W siłowniku o działaniu odwrotnym sprężyna zwrotna jest umieszczona nad membraną, a ciśnienie sterujące jest podawane pod membranę. Wobec tego ciśnienie sterujące powoduje wciąganie trzpienia, podczas gdy w rozwiązaniu przedstawionym na rys. 4.9a ciśnienie sterujące powoduje wysuwanie trzpienia.

W przypadku zaniku ciśnienia sterującego trzpień siłownika o działaniu prostym przyjmie położenie krańcowe górne, a trzpień siłownika o działaniu odwrotnym położenie krańcowe dolne. Zależnie od zastosowanego zaworu oznacza to, że przy zaniku ciśnienia zawór zostaje zamknięty albo otwarty. Dla instalacji technologicznej nie jest oczywiście obojętne, czy awaria zasilania siłownika powoduje otwarcie czy zamknięcie zaworu.

Dla zapewnienia sterowania zaworu w przypadku awarii, siłowniki są wyposażone w napędy umożliwiające ręczne przedstawianie zaworu. Napęd taki nie jest pokazany na rys. 4.9.

·4.10· W jaki sposób zapewnia się zgodność położenia trzpienia siłownika z wartością sygnału sterującego?

Charakterystyka statyczna siłownika membranowego, czyli zależność przesunięcia trzpienia od ciśnienia sterującego $h = f(p_{ster})$, powinna być liniowa. W rzeczywistości, na skutek działania sił tarcia w siłowniku i w zaworze, z którym siłownik jest połączony, a także sił wywieranych na grzybek zaworu, charakterystyka statyczna siłownika ulega zniekształceniu. Dla zorientowania się, jakie są zależności pomiędzy wartościami sił zewnętrznych działających na trzpień siłownika a odchyleniami jego charakterystyki rzeczywistej od charakterystyki liniowej, zastanówmy się, jakie siły działają w siłownikach membranowych. Rozważmy dla przykładu jeden z siłowników produkcji zakładów MERA-POLNA. Zakłady te produkują typoszereg siłowników o następujących parametrach: powierzchnia czynna membrany 290 ... 2129 cm², skok znamionowy wrzeciona 12,7 ... 101,6 mm, ciśnienie sterujące 20 ... 100 kPa (0,2 ... 1 kg/cm²) lub 40 ... 200 kPa (0,4 ... 2 kg/cm²). Siłownik nr kat. 37-13 ma powierzchnię czynną membrany 678 cm²



4.10. Siłownik membranowy z ustawnikiem pozycyjnym

1 — mieszek sprężysty; 2 — dysza; 3 — przysłona; 4 — wzmacniacz mocy; 5 — siłownik; 6 — dźwignia; 7 — sprężyna

i skok wrzeciona 25,4 mm. Przy ciśnieniu sterującym 20 ... 100 kPa (0,2 ... 1 kG/cm²) na membranę działa siła 1360 ... 6800 N. Jeżeli chcemy korzystać z całego zakresu ruchu wrzeciona, zakres naciągu sprężyny powinien być taki, jaki zakres zmian sił działających na membranę. Współczynnik naciągu sprężyny równy ok. 200 N/mm umożliwia spełnienie tych wymagań. Oznacza to, że siła tarcia 200 N spowoduje ok. 1-milimetrowy (czyli ok. 4% zakresu) błąd ustawienia wrzeciona. Jeżeli do sterowania tego siłownika użyjemy ciśnienia 40 ... 200 kPa (0,4 ... 2 kG/cm²), to równocześnie trzeba zastosować dwukrotnie sztywniejszą sprężynę, o współczynniku naciągu równym ok. 400 N/mm. Uzyskamy wówczas dwukrotne zmniejszenie błędu ustawienia wrzeciona.

Radykalną poprawę właściwości siłownika uzyskuje się wyposażając go w tzw. ustawnik pozycyjny. Na rysunku 4.10 przedstawiono schemat połączenia ustawnika pozycyjnego z siłownikiem membranowym; w ustawniku działające na podobnej zasadzie można wyposażać również siłowniki tłokowe.

Zmiana ciśnienia sterującego p_{ster} powoduje zmianę ugięcia mieszk sprężystego 1 i zmianę odległości między dyszą 2 i przysłoną 3. To z kolei powoduje zmianę ciśnienia kaskadowego p_k . Ciśnienie kaskadowe po wzmocnieniu we wzmacniaczu mocy 4 działa na membranę siłownika 5 powodując przesunięcie trzpienia siłownika. Z trzpieniem siłownika jest połączona dźwignia 6 naciskająca na sprężynę 7. Przez

dźwignię 6 i sprężynę 7 jest zrealizowane sprzężenie zwrotne zapewniające dużą dokładność ustawienia trzpienia siłownika. Dzięki działaniu sprzężenia zwrotnego ciśnienie p_{ster} przyjmuje takie wartości, przy których położenie trzpienia siłownika jest zgodne z wartością sygnału p_{ster} , niezależnie od działania czynników zewnętrznych. Dla sprawdzenia, czy tak jest rzeczywiście, założmy, że pojawiła się siła zewnętrzna przesuwająca np. trzpień do góry. Ruch trzpienia wywoła wówczas zmianę położenia przystony w kierunku zmniejszenia jej odległości od dyszy. W konsekwencji nastąpi wzrost ciśnienia p_{ster} i zrównoważenia siły działającej na trzpień.

Stosując ustawnik można usunąć z siłownika sprężynę, a dla uzyskania ruchu powrotnego doprowadzić na przeciwną stronę membrany niż ciśnienie sterujące ciśnienie o stałej wartości, tzw. ciśnienie **podparcia**. W siłownikach bezsprężynowych siła użyteczna uzyskiwana na trzpieniu siłownika jest kilkakrotnie większa niż w siłownikach sprężynowych tych samych rozmiarów.

·4.11· Czym jest ograniczone stosowanie membranowych siłowników pneumatycznych i jakie inne konstrukcje siłowników pneumatycznych można stosować zamiast membranowych?

Siłowniki membranowe wszystkich typów mają wspólną wadę — ograniczony skok trzpienia. Praktycznie bardzo rzadko spotyka się siłowniki membranowe o skoku większym niż 100 mm. Dla większych przesunięć stosuje się siłowniki tłokowe. Produkuje się siłowniki tłokowe pneumatyczne działania jednostronnego (ze sprężyną powrotną) lub działania dwustronnego. Częściej stosuje się siłowniki działania dwustronnego, ponieważ umożliwiają pełne wykorzystanie ciśnienia roboczego dla uzyskiwania siły użytecznej. Siłowniki tłokowe z reguły są wyposażone w ustawniki pozycyjne (w siłownikach działania dwustronnego jest to konieczne). Zawsze też pracują przy podwyższonym ciśnieniu — zwykle przy 600 kPa (6 kG/cm²).

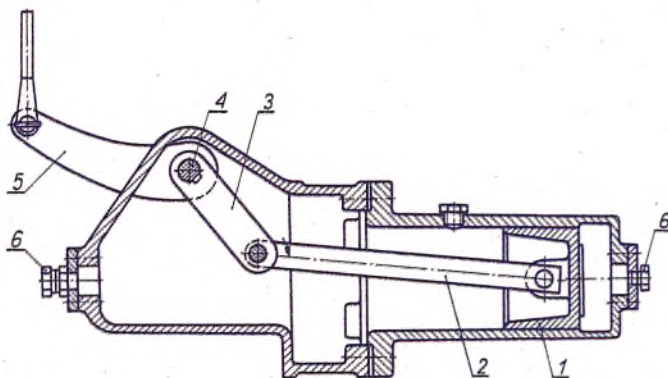
·4.12· Jakiego typu siłowniki hydrauliczne stosuje się w układach regulacji automatycznej?

Siłowniki hydrauliczne mogą być **tłokowe** lub **obrotowe**. W automatyce przemysłowej używa się siłowników tłokowych. Są one budowane jako siłowniki **proste i korbowe**. Na rysunku 4.12 przedstawiono siłownik korbowy. Ruch prostoliniowy tłoka 1 jest przenoszony za pośrednictwem korbowa 2 na korbę 3 i zamieniany na ruch obrotowy korby. Obrót korby 3 umieszczonej wewnątrz szczelnej obudowy siłownika powoduje obrót osi 4 i korby zewnętrznej 5. Wielkością wyjściową siłownika jest obrót korby zewnętrznej 5. Moment użyteczny uzyskiwany na korbie

$$M = F r \cos \alpha$$

4.12. Hydrauliczny siłownik korbowy

1 — tłok; 2 — korbówód; 3 — korba; 4 — oś; 5 — korba zewnętrzna; 6 — doprowadzenia ciśnieniowych przewodów sterujących



przy czym: $F = Ap_{zas}$ (A — powierzchnia tłoka); r — długość korby 3 siłownika; α — kąt między korbówodem a korba.

Kąt α zwykle zmienia się w granicach $45^\circ \dots 135^\circ$, co powoduje zmiany momentu w granicach $70 \dots 100\%$.

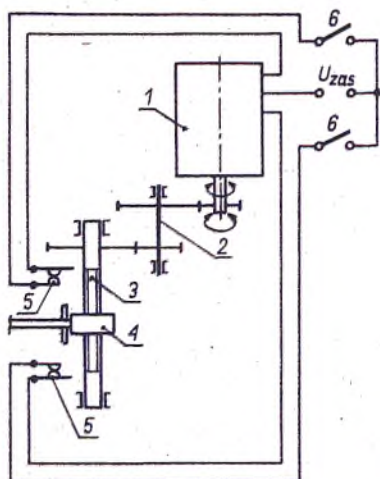
Przedstawiony siłownik jest siłownikiem działania dwustronnego; zależnie od tego, w którą stronę ma być przesunięty tłok, olej pod ciśnieniem jest doprowadzany do komory po lewej lub po prawej stronie tłoka. Siłowniki tłokowe można wykonywać również jako siłowniki działania jednostronne, jednak rozwiązania takie stosuje się bardzo rzadko.

4.13. Jakiego rodzaju elementy napędowe stosuje się w siłownikach elektrycznych?

Produkowane siłowniki elektryczne dzieli się na **silnikowe i elektromagnetyczne**. Jako elementy napędowe w siłownikach elektrycznych silnikowych stosuje się wirujące silniki elektryczne, najczęściej indukcyjne (dwufazowe przy mocach rzędu kilkudziesięciu watów i trójfazowe przy mocach większych). W siłownikach elektromagnetycznych elementami napędowymi są elektromagnesy.

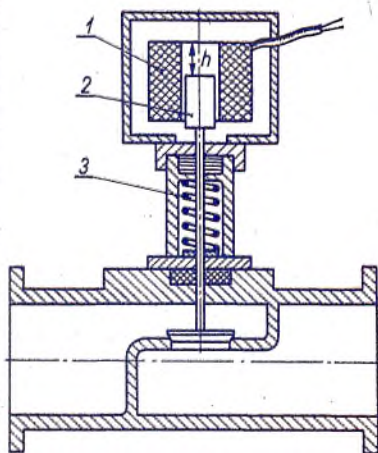
Silniki elektryczne charakteryzują się dużą prędkością obrotową i małym momentem; natomiast do przestawiania zaworów i przepustnic niezbędne są dość duże siły i niewielkie prędkości obrotowe. Wynika stąd konieczność stosowania przekładni mechanicznych zmniejszających prędkość obrotową i zwiększających moment. Jeżeli silnik obraca korbę przez przekładnię zębatą, otrzymujemy **siłownik korbowy**, jeżeli po zredukowaniu prędkości obrotowej silnika ruch obrotowy jest zamieniany na ruch liniowy za pomocą przekładni śrubowej, otrzymujemy **siłownik liniowy**.

Schemat budowy siłownika liniowego przedstawiono na rys. 4.13A. Silnik pracuje systemem załączony-wyłączony. Prędkość obrotowa silnika jest zawsze taka sama, kierunek natomiast zależy od tego, która z dwu par styków 6 jest zwarta. Po załączeniu silnika trzpień wyjściowy siłownika przemieszcza się ze stałą prędkością tak długo, jak długo silnik nie zostanie wyłączony. W siłowniku są zainstalowane wyłącz-



4.13A. Schemat budowy i działania siłownika elektrycznego silnikowego

1 — silnik; 2 — przekładnia zębata; 3 — śruba; 4 — nakrętka; 5 — wyłączniki krańcowe; 6 — przełączniki załączające silnik



•4.13B• Siłownik elektromagnetyczny
1 — uzwojenie elektromagnesu; 2 — rdzeń;
3 — sprężyna zwrotna

niki krańcowe 5. Ich zadaniem jest odłączenie silnika zanim zostanie on zahamowany na skutek dojścia do konstrukcyjnych ograniczeń w siłowniku lub urządzeniach współpracujących, np. silnik powinien być wyłączony po całkowitym zamknięciu zaworu. Pozostawienie włączonego silnika w stanie zahamowanym jest bowiem niewskazane, a dla wielu silników zabronione, ze względu na możliwość ich uszkodzenia.

Siłowniki elektromagnetyczne stosuje się do sterowania zaworów (otwarty-zamknięty) o niewielkich średnicach znamionowych. Dość często siłowniki elektromagnetyczne stosuje się jako zawory odcinające. Budowę siłownika elektromagnetycznego pokazuje rys. 4.13B. Przepływ prądu przez uzwojenie 1 elektromagnesu powoduje powstanie siły wciągającej rdzeń 2 i natychmiastowe wciągnięcie rdzenia do góry, co powoduje całkowite otwarcie zaworu. Po wyłączeniu prądu zasilającego elektromagnes zawór zamyka sprężyna 3. Stosowane są również konstrukcje, w których załączenie elektromagnesu powoduje zamknięcie zaworu, a wyłączenie — otwarcie.

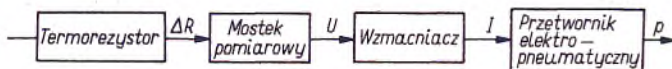
•4.14. Co to są układy pomiarowe?

Układy pomiarowe są to zespoły urządzeń dostarczających informacji o stanie regulowanego procesu technologicznego. Źródłem sygnału w układzie pomiarowym jest **czujnik pomiarowy**, tzn. ta część układu, na którą bezpośrednio działa wielkość mierzona. Sygnał wyjściowy czujnika tylko niekiedy nadaje się do bezpośredniego wykorzystania w układzie regulacji, na ogół trzeba go przekształcić do postaci zgodnej z wymaganiami stawianymi przez regulatory. Układy pośredniczące przetwarzające sygnały wyjściowe czujników na sygnały standardowe nazywamy **przetwornikami pomiarowymi**. Są one niezbędne, ponieważ znaczna różnorodność parametrów mierzonych w różnego rodzaju obiektach powoduje, że sygnały wyjściowe czujników też są bardzo różne. Systemy blokowe aparatury regulacyjnej (patrz pyt. 4.3) zawierają przetworniki pomiarowe przeznaczone do współpracy z najważniejszymi typami czujników. Istnieją również przetworniki pomiarowe przekształcające standardowe sygnały elektryczne na pneumatyczne i odwrotnie, przetworniki cyfrowo-analogowe, przetworniki analogowo-cyfrowe i in.

Sygnały wyjściowe czujników są przetwarzane często kilkakrotnie. Na rysunku 4.14 przedstawiono schemat blokowy układu z czujnikiem rezystancyjnym do pomiaru temperatury i wytwarzającego standardowy pneumatyczny sygnał wyjściowy.

Pomiaru niektórych wielkości fizycznych rzadziej spotyka-

•4.14• Schemat blokowy układu pomiaru temperatury z czujnikiem elektrycznym i pneumatycznym sygnałem wyjściowym



nych lub wymagających specyficznych metod pomiarowych nie można realizować przy użyciu standardowego układu czujnik-przetwornik. Stosuje się wtedy specjalne aparaty nie wchodzące w skład systemów blokowych, dopasowując ewentualnie ich sygnały wyjściowe za pomocą przetworników.

4.15. Jakie właściwości urządzeń pomiarowych są ważne przy stosowaniu ich w układach regulacji automatycznej?

Właściwości każdego urządzenia określa jego **charakterystyka statyczna i dynamiczna**. Charakterystyka statyczna rzeczywistego urządzenia pomiarowego

$$x_{wy} = f(x_{we})$$

różni się od charakterystyki idealnej

$$x_{wy}^* = f(x_{we})$$

i w związku z tym wynik pomiaru jest obciążony błędem którego wartość bezwzględna

$$\Delta x_{wy} = x_{wy} - x_{wy}^*$$

i wartość względna

$$\delta = \frac{\Delta x_{wy}}{x_{wy \max} - x_{wy \min}}$$

Maksymalna wartość względnego błędu pomiaru określonego przy zmienionych warunkach pracy nazywana jest błędem **podstawowym** albo **granicznym** urządzenia.

Błąd podstawowy, określony dla danego urządzenia nazywamy jego **klasą dokładności**. Poszczególne egzemplarze urządzenia mogą posiadać różne błędy podstawowe, nie przekraczające jednak klasy dokładności. Zależnie od potrzeb można wyróżnić inne właściwości statyczne urządzenia takie jak **histereza, próg czułości, pełzanie zera, liniowość, niestalość charakterystyki** itp.

Oprócz błędu podstawowego urządzenie może wykazywać **błędy dodatkowe**, wywołane przez niedotrzymanie znamionowych warunków pracy takich jak warunki zasilania, temperatura otoczenia, wilgotność itp.

Zakresem pomiarowym urządzenia nazywamy zakres zmian wielkości wejściowej, w którym pracuje ono z zachowaniem określonej klasy dokładności. Należy podkreślić, że przy szeregowymłączeniu czujnika i przetworników, ich błędy mogą się sumować. Dla określenia klasy powstałego przez takie połączenie urządzenia pomiarowego należy więc dodać do siebie podane w procentach klasy czujnika i przetworników.

Właściwości dynamiczne określa się przez podanie **częstotliwości granicznej**, przy której występuje pewne, dopuszczalne zmniejszenie wzmocnienia (czyli dopuszczalna war-

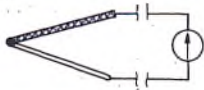
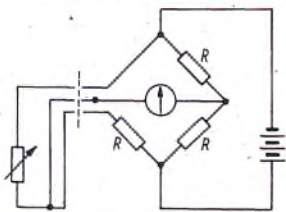
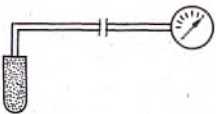
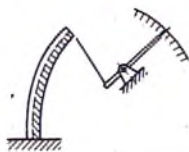
tość błędu dodatkowego). W większości przypadków częstotliwości graniczne urządzeń pomiarowych są dużo większe niż częstotliwość zmian wielkości mierzonych (wielkości wyjściowych obiektów). W tych przypadkach można pominąć istnienie dodatkowego błędu dynamicznego.

Urządzenie pomiarowe włącza się do układu regulacji tak jak pokazano na rys. 4.1. Porównywanie wartości wielkości regulowanej z wartością zadaną odbywa się za urządzeniem pomiarowym, na wejściu regulatora. Dokładność regulacji uzyskiwana w całym układzie regulacji nie może być większa niż dokładność urządzenia pomiarowego. Dokładność regulacji jest zatem ograniczona wartością błędu podstawowego oraz wartościami błędów dodatkowych (statycznych i dynamicznych) urządzenia pomiarowego.

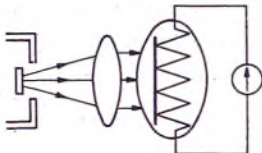
·4.16· Jakie przyrządy stosuje się do pomiaru temperatury?

Najważniejsze rodzaje przyrządów pomiarowych używanych w układach automatyki do pomiaru temperatury są zestawione w tabl. 4.16.

4.16

Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres pomiaru	Dokładność
Temperatura		Termoelement Napięcie	$-200 \dots 1400^{\circ}\text{C}$	$0,1 \dots 0,5^{\circ}\text{C}$
		Czujnik rezystancyjny Zmiana rezystancji	$-258 \dots 1000^{\circ}\text{C}$	$0,1 \dots 1^{\circ}\text{C}$
		Czujnik manometryczny Zmiana ciśnienia	$-240 \dots 500^{\circ}\text{C}$	$0,1 \dots 1\%$
		Czujnik bimetaliczny Przesunięcie	$-180 \dots 450^{\circ}\text{C}$	$0,5 \dots 2\%$

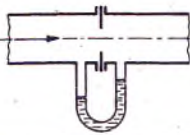
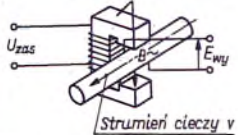
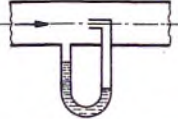
4.16 (cd.)

Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres pomiaru	Dokładność
Temperatura		Czujnik cieczowy Wysokość słupa cieczy	$-100 \dots 600^{\circ}\text{C}$	$0,01 \dots 0,1^{\circ}\text{C}$
		Pirometr Napięcie lub natężenie prądu stałego	$400 \dots 3500^{\circ}\text{C}$	ok. 2%

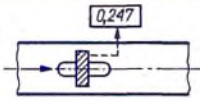
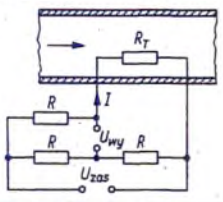
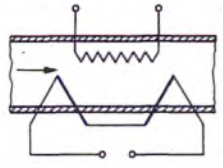
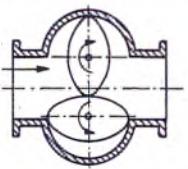
4.17. Jakie przyrządy stosuje się do pomiaru przepływu?

Najważniejsze rodzaje przyrządów pomiarowych używanych w układach automatyki do pomiaru przepływu są zestawione w tabl. 4.17.

4.17

Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres zastosowań	Dokładność
Natężenie przepływu		Zwężka Różnica ciśnień	Bardzo szeroki	Zależna od warunków pracy i właściwości czynnika
		Rotametr Położenie pływak	Stosowany dla średnic przewodów $1 \dots 30 \text{ cm}$	$0,5 \dots 2\%$
		Czujnik indukcyjny Napięcie	Rury dowolnych średnic, prędkość przepływu $10 \dots 300 \text{ cm/s}$	$0,5 \dots 1\%$
		Rurka spiętrzająca Różnica ciśnień	Ograniczony	Niewielka

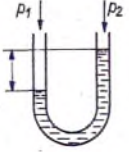
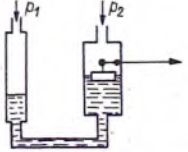
4.17 (cd.)

Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres zastosowań	Dokładność
Natężenie przepływu		Turbinka lub wiatraczek	Stosowany przy średnicach rur do kilkudziesięciu cm	0,1...1%
		Termoanemometr Napięcie	0...8000 m/min	2...5%
		Kalorymetryczny Napięcie	Małe prędkości przepływu	2...5%
		Czujnik dozujący	0,3...70 m/h	ok. 0,5%

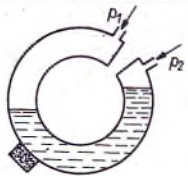
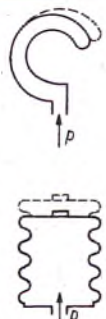
.4.18. Jakie przyrządy stosuje się do pomiaru ciśnień?

Przyrządy do pomiaru ciśnień są zestawione w tabl. 4.18.

4.18

Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres pomiaru	Dokładność
Ciśnienie		Manometr cieczowy Różnica poziomów	0...100 kPa	0,5 mm
		Czujnik cieczowy pływakowy Przesunięcie	0...50 kPa	ok. 1%

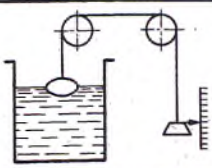
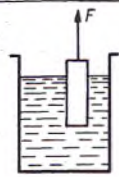
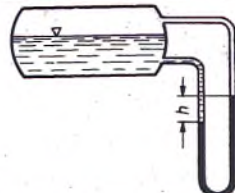
4.18 (c.d.)

Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres pomiaru	Dokładność
Ciśnienie		Waga pierścieniowa Kąt obrotu	0...25 kPa	ok. 1,5%
		Elementy sprężyste Przesunięcie	0,5...10 ⁵ kPa	1...1,5%

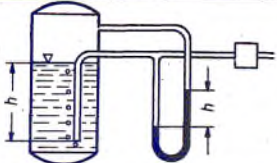
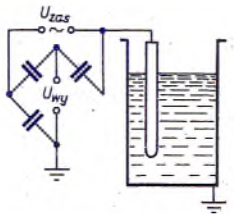
4.19. Jakie przyrządy stosuje się do pomiaru poziomu cieczy?

Przyrządy do pomiaru cieczy są zestawione w tabl. 4.19.

4.19

Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres pomiaru	Dokładność
Poziom cieczy		Pływak Przesunięcie	Nie ograniczony	do ok. 2%
		Pływak wypornościowy Siła	Równy długości pływaka	1...2%
		Czujnik manometryczny	Ograniczony zakresem manometru	ok. 1%

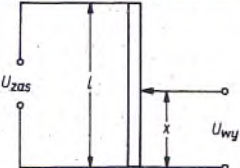
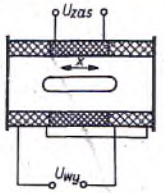
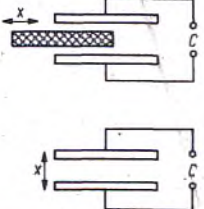
4.19 (cd.)

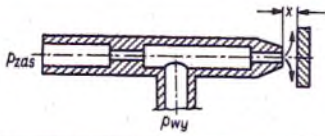
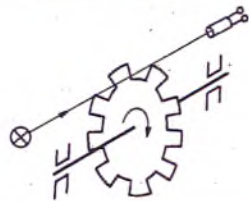
Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres pomiaru	Dokładność
Poziom cieczy		Czujnik manometryczny Ciśnienie	Jak wyżej	ok. 1%
		Czujnik pojemnościowy Zmiana pojemności	Nie ograniczony	ok. 2,5%

4.20. Jakie przyrządy stosuje się do pomiaru przesunięć liniowych i kątowych?

Przyrządy do pomiaru przesunięć liniowych i kątowych są zestawione w tabl. 4.20.

4.20

Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres pomiaru	Dokładność
Przesunięcia liniowe i kątowe		Potencjometr Napięcie	Kilkadziesiąt cm	ok. 0,1%
		Czujnik indukcyjny transformatorowy Napięcie	Od kilku mikrometrów do kilku cm	ok. 0,1%
		Czujnik pojemnościowy Zmiana pojemności	Jak wyżej	ok. 0,1%

Wielkość mierzona	Zasada działania (schemat funkcjonalny)	Rodzaj czujnika i sygnał wyjściowy	Zakres pomiaru	Dokładność
Przesunięcia liniowe i kątowe		Czujnik pneumatyczny Ciśnienie	Od kilkunastu mikrometrów do kilku mm	ok. 0,5%
		Czujnik impulsowy Liczba impulsów	Ograniczony pojemnością licznika	1 impuls

4.21. Co to jest regulator?

Regulatorem nazywa się w automatyce aparat, który wytwarza sygnał sterujący proces technologiczny w sposób zapewniający przebieg procesu zgodny z przebiegiem pożądanym. W regulatorze następuje:

— Porównanie aktualnej wartości zmiennej kontrolowanej z wartością zadaną tej zmiennej (określenie wartości uchybu regulacji).

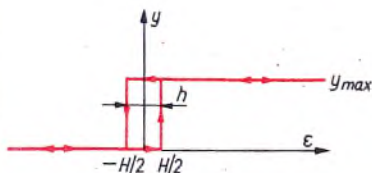
— Wytworzenie sygnału wyjściowego o wartości zależnej od wartości uchybu regulacji, czasu występowania uchybu i szybkości jego zmian.

Sygnał wyjściowy regulatora powinien posiadać moc i postać umożliwiającą uruchamianie urządzeń wykonawczych. Sygnał wartości zadanej (porównywany z sygnałem wartości bieżącej) jest wytwarzany z reguły także w regulatorze przez tzw. **nadajnik wartości zadanej**. Nadajnik wartości zadanej jest częścią tzw. **stacyjki sterowania ręcznego**, stanowiącej przeważnie wydzielony konstrukcyjnie zespół regulatora. W skład stacyjek sterowania ręcznego wchodzi także: urządzenia do ręcznego sterowania obiektem, mierniki mierzące wartość wielkości istotnych dla procesu regulacji (wielkość regulowana, wartość zadana, sygnał wyjściowy regulatora) oraz przełączniki rodzaju pracy regulatora. Podstawowymi rodzajami pracy są: „praca automatyczna” i „praca ręczna”; odpowiada im: sterowanie obiektu sygnałem wyjściowym z regulatora (praca automatyczna) i z nadajnika sygnału sterowanego przez człowieka (praca ręczna). W niektórych regulatorach możliwe są jeszcze inne rodzaje pracy, jak np. praca w układzie kaskadowym, współpraca z komputerem i inne.

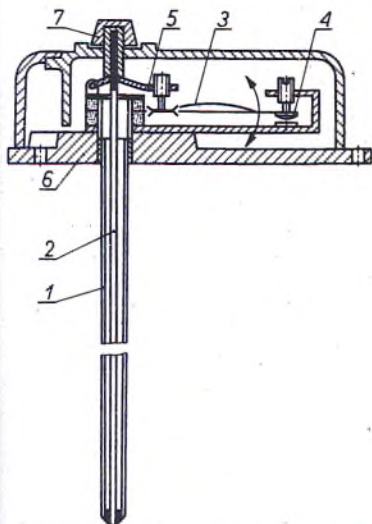
Z punktu widzenia teorii regulacji, regulator jest szeregowym członem korekcyjnym (por. pyt. 3.26).

•4.22• Jakie są rodzaje regulatorów?

Istnieje kilka kryteriów podziału regulatorów. Przy podziale według sposobu dostarczania energii zasilającej, potrzebnej do oddziaływania na obiekt (czyli do przestawiania elementu nastawczego), wyróżniamy regulatory nie korzystające z energii pomocniczej (**bezpośredniego działania**) i regulatory korzystające z energii pomocniczej. Regulatory bezpośredniego działania pobierają potrzebną do działania energię z urządzenia pomiarowego — są to z reguły proste konstrukcje umożliwiające uzyskiwanie niewielkich dokładności regulacji. Regulatory z energią pomocniczą dzieli się, według nośnika tej energii, na **elektryczne, pneumatyczne i hydrauliczne**. Przy klasyfikacji według postaci sygnału wyjściowego wyróżniamy regulatory o **wyjściu ciągłym**, regulatory **impulsowe** oraz regulatory **dwu- i trójpoleżeniowe**. Wśród regulatorów ciągłych istnieje grupa regulatorów, w których wzmacniacz regulatora jest nieciągły, ale dzięki stosowaniu członu wykonawczego o własnościach całkujących uzyskuje się ciągłość zmian sygnału sterującego — są to regulatory **krokowe**.



•4.23A• Charakterystyka statyczna regulatora dwupołożeniowego

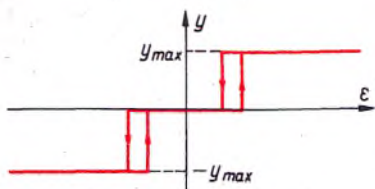


•4.23B• Termoregulator dylatacyjny
1 — tuleja; 2 — pręt; 3 — sprężyna; 4 — styk;
5 — dźwignia; 6 — korpus; 7 — pokrętło do
ustawiania wartości zadanej

•4.23• Co to są regulatory dwu- i trójpoleżeniowe, jakie mają właściwości i gdzie są stosowane?

Regulatory dwupołożeniowe umożliwiają załączanie i wyłączenie dopływu energii lub czynnika do obiektu, zależnie od wartości uchybu regulacji. Widać to dokładnie na charakterystyce statycznej regulatora dwupołożeniowego (rys. 4.23A). Dla wartości uchybu mniejszych od $-H/2$ (wartość H nazywamy szerokością pętli histerezy) sygnał wyjściowy regulatora jest równy zero, dla wartości uchybu większych od $+H/2$ sygnał wyjściowy regulatora przyjmuje wartość maksymalną. Dla wartości uchybu większych od $-H/2$ i mniejszych od $+H/2$ sygnał wyjściowy regulatora jest zerowy lub maksymalny zależnie od tego, czy któraś z tych wartości jest osiągnięta od strony ujemnych czy od strony dodatnich wartości uchybu.

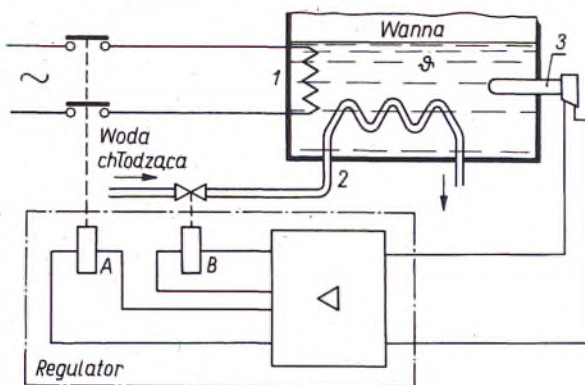
Jako przykład regulatora dwupołożeniowego omówimy regulator temperatury przedstawiony na rys. 4.23B. Czujnikiem w tym regulatorze jest tzw. **termometr dylatacyjny** (rozszerzalnościowy) składający się z tulei 1 i pręta 2, wykonanych z materiałów różniących się znacznie współczynnikami rozszerzalności cieplnej. Przy zmianach temperatury następuje przemieszczanie końca pręta względem końca tulei. Ponieważ tuleja jest przymocowana sztywno do obu-



•4.23C• Charakterystyka statyczna regulatora trójpółożeniowego

•4.23D• Układ regulacji temperatury z regulatorem trójpółożeniowym

1 — grzejnik; 2 — chłodnica



dowy, więc przy zmianach temperatury powstają przesunięcia końca pręta względem obudowy. Przesunięcia te przenoszone są przez dźwignię 5 na sprężynę 3. Zależnie od nacisku wywieranego przez dźwignię na sprężynę zajmuje ona jedno z położów skrajnych, zwierając lub rozwierając styki łączące lub wyłączające zasilanie grzejnika w obiekcie. Do ustawiania wartości zadanej służy pokrętko 7.

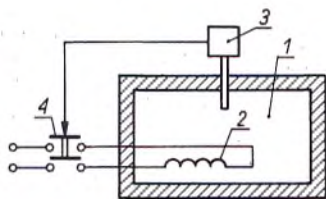
Charakterystykę statyczną regulatora trójpółożeniowego pokazuje rys. 4.23C. Wielkość wyjściowa regulatora może przyjmować wartość maksymalną dodatnią, maksymalną ujemną albo może być równa zero.

Na rysunku 4.23D przedstawiono przykład zastosowania regulatora trójpółożeniowego. Jest to układ stabilizacji temperatury wody w wannie służącej do hartowania wyrobów stalowych. W początkowym okresie pracy wanny temperatura znajdującej się w niej wody jest za niska i trzeba ją podwyższyć. Do podgrzewania wody służy grzejnik włączany w razie potrzeby przez przekąznik A. Jeżeli na skutek zanurzenia w wannie gorących przedmiotów temperatura wody wzrośnie nadmiernie, wówczas przekąznik B powoduje otwieranie zaworu doprowadzającego wodę chłodzącą.

•4.24• Jakie wyniki daje regulacja dwupółożeniowa?

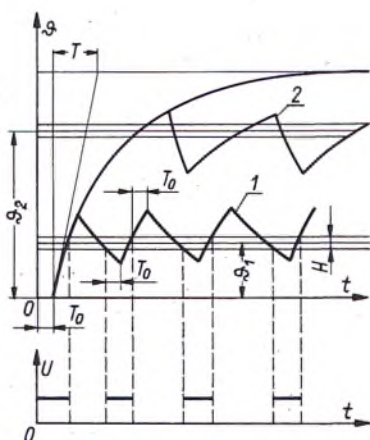
Typowym zastosowaniem regulatorów dwupółożeniowych jest regulacja temperatury. Dla oceny wyników tego typu regulacji zbadamy zjawiska zachodzące w układzie, którego schemat przedstawiono na rys. 4.24A.

Włączenie obiektu do sieci następuje, gdy temperatura obiektu jest niższa od zadanej, więc grzejnik jest włączony. Wobec tego temperatura obiektu wzrasta (rys. 4.24B). Po osiągnięciu temperatury ϑ_2 grzejnik zostaje wyłączony, ale ponieważ w obiekcie istnieje opóźnienie T_0 , temperatura w dalszym ciągu wzrasta. Dopiero po czasie równym T_0 temperatura obiektu zaczyna maleć. Kiedy temperatura zmaleje do wartości ϑ_1 , grzejnik jest ponownie włączony. Jednak, tak jak przy wyłączaniu grzejnika, efekt włączenia zaobserwujemy dopiero po upływie czasu T_0 . Gdy temperatura osiągnie



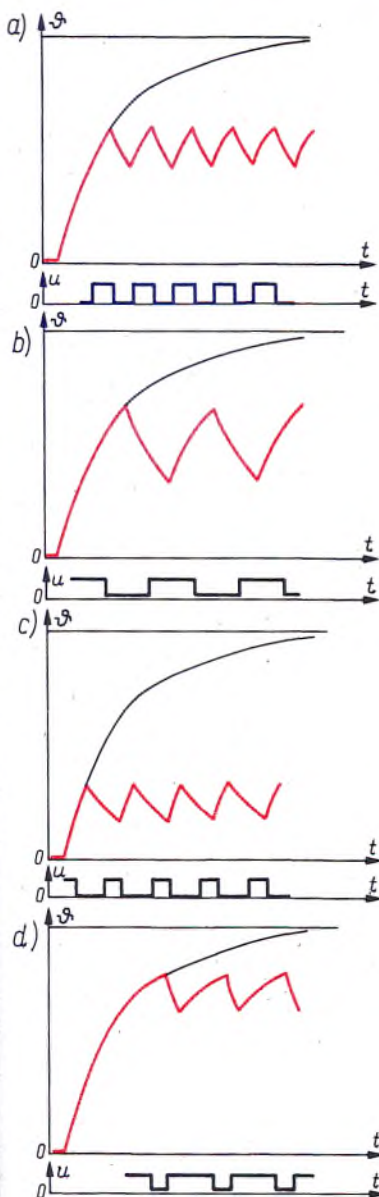
•4.24A• Układ dwupółożeniowej regulacji temperatury

1 — obiekt; 2 — grzejnik; 3 — regulator dwupółożeniowy; 4 — stycznik



•4.24B• Przebiegi wielkości regulowanej w układzie regulacji dwupółożeniowej

1 — przy małej wartości zadanej; 2 — przy dużej wartości zadanej



4.25A. Przebiegi otrzymywane na wyjściu obiektu inercyjnego z opóźnieniem: a, b) przy różnych częstotliwościach; c, d) przy różnych współczynnikach wypełnienia prostokątnego sygnału sterującego

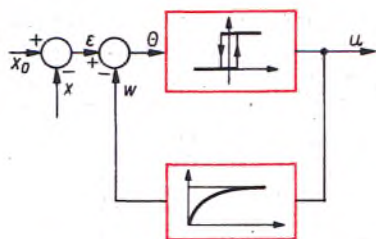
wartość ϑ_2 , grzejnik zostanie wyłączony itd.; cykl będzie się powtarzał. Temperatura oscyluje wokół wartości zadanej. Amplituda oscylacji zależy od szerokości pętli histerezy oraz od stosunku opóźnienia do stałej czasowej obiektu. Rysunek pokazuje także zależność przebiegu temperatury od wartości zadanej (przebiegi 1 i 2).

4.25. Jak można poprawić jakość regulacji dwupołożeniowej?

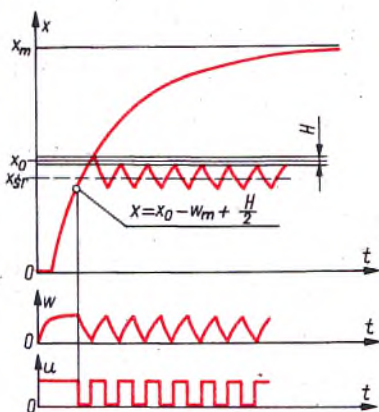
Największą niedogodnością regulacji dwupołożeniowej stanowią ciągłe oscylacje wielkości regulowanej. Oscylacje te wynikają z zastosowania elementu dwupołożeniowego do sterowania pracy obiektu i nie można ich wyeliminować nie rezygnując ze stosowania elementu tego typu. Próby poprawy jakości regulacji dwupołożeniowej polegają więc na działaniu w kierunku zmniejszania amplitudy oscylacji.

Amplituda oscylacji w układzie przedstawionym na rys. 4.24B zależy od stosunku czasów T_0/T obiektu. Od tych samych parametrów obiektu zależy częstotliwość włączeń grzejnika w tym układzie. Częstotliwość włączeń i amplituda oscylacji są ze sobą ściśle związane. Jeżeli bowiem okres włączenia jest długi, to towarzyszy mu duża zmiana temperatury. To samo dotyczy czasu wyłączenia grzejnika. Na rysunku 4.25A przedstawiono przebiegi zmian temperatury otrzymane przy sterowaniu obiektu sygnałami o różnej częstotliwości i różnym wypełnieniu. Jak widać, zwiększenie częstotliwości przebiegu sterującego powoduje zmniejszenie amplitudy oscylacji. Wypełnienie (stosunek czasu włączenia do czasu wyłączenia) decyduje natomiast o średniej wartości temperatury w obiekcie. Z przedstawionych przebiegów wynika więc, że dla zmniejszenia amplitudy oscylacji należy zwiększyć częstotliwość włączeń grzejnika. Można to uzyskać wprowadzając do układu regulacji ujemne sprzężenie zwrotne poprzez człon inercyjny (rys. 4.25B). Powstaje w ten sposób jak gdyby wewnętrzny układ regulacji dwupołożeniowej, dla którego wartością zadaną jest sygnał uchybu z głównego układu regulacji. W układzie tym powstają oscylacje o częstotliwości zależnej od szerokości pętli histerezy elementu dwupołożeniowego oraz od stałej czasowej układu sprzężenia zwrotnego. Dobierając odpowiednio stałą czasową układu sprzężenia zwrotnego można uzyskać potrzebne zwiększenie częstotliwości przełączeń elementu dwupołożeniowego sterującego obiektem.

Przebiegi czasowe powstające w układzie regulacji dwupołożeniowej z korekcją przedstawiono na rys. 4.25C. Charakter drgań, jakie powstają w tym układzie, zależy w dużym stopniu od parametrów układu sprzężenia zwrotnego. Gdy szybkość zmian sygnału sprzężenia zwrotnego jest dużo większa niż szybkość zmian sygnału wyjściowego z obiektu, wówczas częstotliwość przełączeń w układzie zależy tylko od szerokości pętli histerezy regulatora i stałej czasowej układu sprzężenia zwrotnego.



•4.25B• Schemat blokowy regulatora dwupołożeniowego z korekcją przez sprzężenie zwrotne

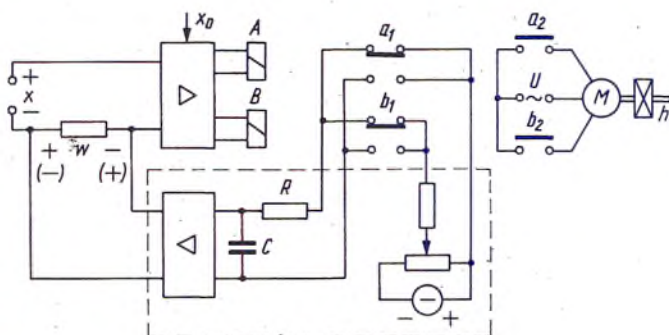


•4.25C• Przebiegi czasowe w układzie regulacji dwupołożeniowej z korekcją

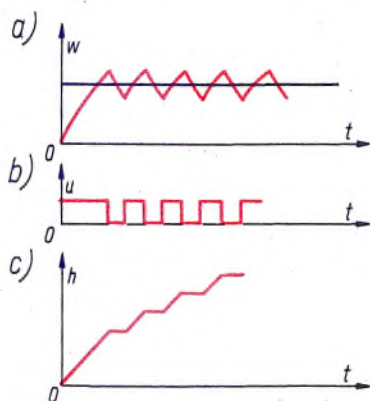
•4.26• Co to jest regulator krokowy?

Regulator krokowy otrzymuje się wprowadzając do regulatora trójpoleźniowego układ ujemnego inercyjnego sprzężenia zwrotnego. Przykład takiego rozwiązania przedstawia rys. 4.26A. Jeżeli różnica między napięciem x (reprezentującym wielkość regulowaną) a wartością zadaną x_0 jest większa od strefy nieczułości regulatora (patrz charakterystyka statyczna regulatora, rys. 4.23C), wówczas działa jeden z przełączników, A lub B . Jeżeli wartość x jest mniejsza od x_0 , tzn. jeżeli uchyb regulacji jest dodatni, działa przełącznik B . Zadziałanie jednego z tych przełączników powoduje podanie (odpowiednio przez styki a_1 lub b_1), napięcia stałego na wejście układu RC . To z kolei powoduje pojawienie się na wyjściu przystawki sprzężenia zwrotnego napięcia w narastającego zgodnie ze stałą czasową układu RC . Polaryzacja sygnału w zależy od tego, który z przełączników zadziałał i zawsze jest przeciwna do znaku uchybu regulacji. Polaryzację sygnału w zaznaczono na rys. 4.26A, — zadziałaniu przełącznika A odpowiadają oznaczenia bez nawiasów, zadziałaniu przełącznika B — oznaczenia w nawiasach.

Jeżeli wartość bezwzględna uchybu jest mniejsza niż suma strefy nieczułości i maksymalnej wartości sygnału w , to po pewnym czasie od zadziałania jednego z przełączników różnica pomiędzy uchybem i sygnałem w stanie się mniejsza od strefy nieczułości (w rozważaniach tych przyjmujemy, że szybkość zmian sygnału x jest pomijalnie mała w porównaniu z szybkością zmian sygnału w) i przełącznik zwolni. Zwolnienie przełącznika spowoduje odłączenie napięcia od członu RC i rozładowanie się kondensatora C . Po pewnym czasie, zależnym od stałej czasowej rozładowania kondensatora, gdy wartość sygnału w zmniejszy się o wartość równą szerokości pętli histerezy, przełącznik ponownie zadziała i wartość sygnału w zacznie wzrastać. Gdy w zwiększy się na tyle, że różnica między nim a wartością uchybu stanie się mniejsza od strefy nieczułości, przełącznik zwolni i sygnał w zacznie maleć. Opisany proces będzie przebiegał cyklicznie i w rezultacie otrzymamy kolejne włączanie i wyłączenie jednego z przełączników regulatora. Stosunek czasu włączenia do



•4.26A• Schemat połączenia regulatora trójpoleźniowego RK-2 z przystawką sprzężenia zwrotnego RF-6



4.26B. Przebiegi w układzie regulatora trójpółożeniowego ze sprzężeniem zwrotnym sterującym silnik, po pojawieniu się uchybu: a) sygnał na wyjściu członu sprzężenia zwrotnego; b) stan przełącznika; c) położenie wału silnika

czasu wyłączenia przełączników zależy od wartości uchybu. W układzie przedstawionym na rys. 4.26A zadziałanie przełącznika *A* lub *B* włącza poprzez styki a_2 lub b_2 silnik przestawiający zawór, co powoduje zwiększenie wartości x , gdy działa przełącznik *A*, oraz zmniejszenie wartości x , gdy działa przełącznik *B*. Na rysunku 4.26B przedstawiono przebiegi sygnału w , stanu przełącznika (zależnie od znaku uchybu może to być przełącznik *A* lub *B*) oraz przesunięcia grzybka zaworu h sterowanego silnikiem *M*. Na podstawie tych przebiegów można stwierdzić, że:

- pojawienie się uchybu powoduje powstanie w regulatorze oscylacji;

- stosunek czasu włączenia do czasu wyłączenia silnika zależy od wartości uchybu;

- wykres zmiany położenia zaworu, spowodowanej skokową zmianą wartości uchybu, jest zbliżony do odpowiedzi skokowej regulatora typu PI.

Regulatory działające tak jak omówiony układ nazywane są regulatorami krokowymi, ponieważ silnik wykonawczy wykonuje jak gdyby „kroki” o długości zależnej od wartości uchybu. Omówiony układ ma właściwości zbliżone do regulatora typu PI. Budowane są także regulatory (np. regulator typu ARK-21 produkowany przez MERA-ELPO) o właściwościach regulatorów typu PID. Jeżeli obiekt ma dużą stałą czasową, a do takich obiektów należą obiekty cieplne, wówczas wyniki uzyskiwane przy użyciu regulatorów krokowych nie różnią się od wyników uzyskiwanych za pomocą regulatorów ciągłych.

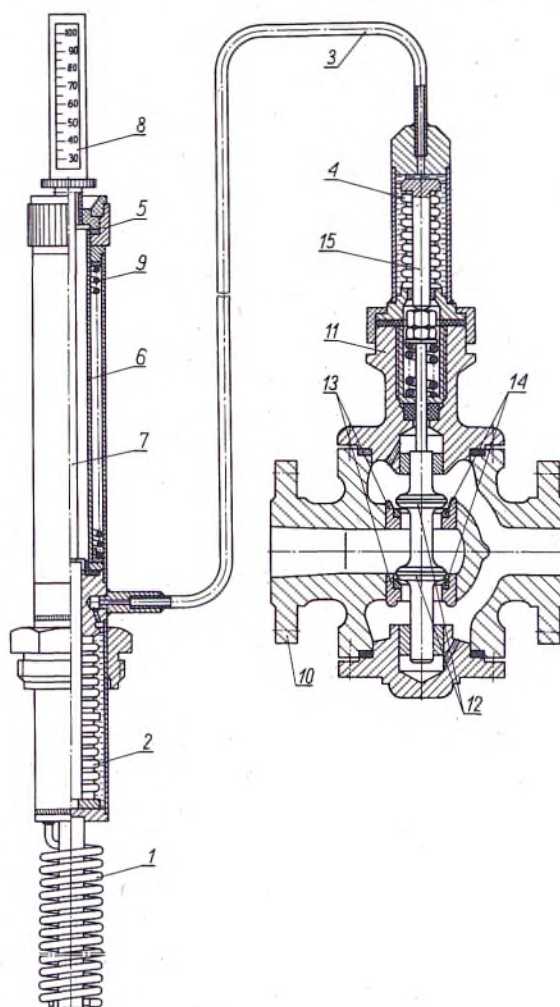
4.27. Co to są regulatory bezpośredniego działania i kiedy się je stosuje?

Regulatorami bezpośredniego działania nazywamy regulatory, które energię potrzebną do przestawiania zaworu nastawczego czerpią z procesu regulowanego za pośrednictwem czujnika pomiarowego. Nie wymagają one stosowania dodatkowych źródeł zasilania, np. sprężarki lub pompy olejowej. Zaletą regulatorów bezpośredniego działania jest prosta i zwarta budowa oraz wynikająca stąd stosunkowo mała cena i duża niezawodność działania. Jednak zakres ich stosowania ogranicza się do regulacji stałowartościowej o małej dokładności. Wynika to z faktu, że w regulatorach tych nie można uzyskać działania typu PI oraz typu PID, jak również nie można uzyskać dużych mocy do napędu zaworów. Regulatory bezpośredniego działania są najczęściej regulatorami proporcjonalnymi. Konieczność uzyskania z czujnika mocy wystarczającej do przestawiania zaworu nastawczego ogranicza zakres stosowania regulatorów bezpośredniego działania do układów regulacji takich wielkości, przy których pomiarze można stosować czujniki z sygnałem wyjściowym w postaci ciśnienia bądź przesunięcia. Do wielkości takich należą: ciśnienie, temperatura, natężenie przepływu i poziom cieczy.

·4.28· Jak jest zbudowany typowy regulator bezpośredniego działania?

Jako przykład regulatora bezpośredniego działania omówimy regulator temperatury. Budowę regulatora (typu BTO produkcji zakładów MERA-POLNA, Przemyśl) przedstawia rys. 4.28.

Czujnikiem tego regulatora jest termometr manometryczny. Spirala termometryczna 1, worek przegrzania 2, kapilara 3 i worek wykonawczy 4 są wypełnione cieczą manometryczną. Spirala termometryczna jest zanurzona w czynniku, którego temperatura podlega regulacji. Temperatura cieczy manometrycznej jest więc taka sama jak temperatura obiektu kontrolowanego. Pod wpływem zmian temperatury zmienia się objętość cieczy manometrycznej. Wzrost temperatury, a więc rozszerzanie się cieczy manometrycznej powoduje ściskanie worka wykonawczego 4 i zamykanie zaworu



·4.28· Regulator temperatury bezpośredniego działania typu BTO

1 — spirala termometryczna; 2 — worek przegrzania; 3 — kapilara; 4 — worek wykonawczy; 5 — pokrętko; 6 — nakrętka; 7 — trzpień gwintowany; 8 — wskaźnik; 9 — sprężyna; 10 — korpus zaworu; 11 — dławnica; 12 — grzybek zaworu; 13 — gniazda zaworu; 14 — wkładki teflonowe; 15 — popychacz worka wykonawczego

przez popychacz 15. Otwieranie zaworu przy obniżaniu temperatury odbywa się za pomocą sprężyny.

Jako zawór wykonawczy regulatora zastosowano zawór dwugniazdowy, charakteryzujący się mniejszymi, w porównaniu z zaworami innej konstrukcji, siłami działającymi na grzybek ze strony czynnika znajdującego się w rurociągu. Zawór jest montowany na rurociągu doprowadzającym do obiektu gorącą parę lub ciecz. Czujnik jest montowany w obiekcie. Odległość między czujnikiem i zaworem nie może przekraczać 2 m, ponieważ taka jest długość kapilary łączącej oba te urządzenia.

Do ustawienia wartości zadanej służy pokrętło 5 połączone na stałe z nakrętką 6. Obracając pokrętłem wymuszamy przesunięcie nagwintowanego trzpienia 7. Ponieważ trzpień 7 jest na stałe połączony z dnem mieszka przegrzania 2, więc wsuwanie lub wysuwanie trzpienia jest równoznaczne ze zmianami ugięcia mieszka. Wobec nieściśliwości cieczy manometrycznej, zmiany ugięcia mieszka 2 wymuszają zmiany ugięcia mieszka wykonawczego 4, zależnie od zmian temperatury czynnika. Zmiany długości mieszka są przekazywane na wskaźnik 8, na którym jest podziałka wyskalowana w stopniach Celsjusza. Na podziałce tej można odczytać, jaką ustawiono wartość zadaną temperatury.

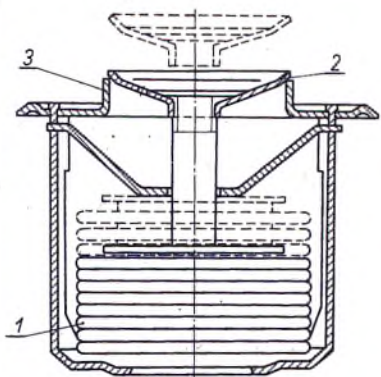
Regulator ma zabezpieczenie przed uszkodzeniem mieszków na skutek przegrzania czujnika. Zakres zmian temperatury regulowanej, powodujący pełne przestawienie zaworu, wynosi zależnie od wykonania 6 ... 8°C. Temperaturę zadaną można ustawiać w zakresie od 30 do 100°C — jest to typowy zakres dla regulatorów przeznaczonych do pracy w urządzeniach grzewczych (np. w wymiennikach ciepła). Regulator ten jest regulatorem proporcjonalnym.

•4.29• Czy termostat samochodowy jest regulatorem?

Budowę termostatu samochodowego, służącego do regulacji temperatury wody w układzie chłodzenia silnika samochodowego przedstawiono na rys. 4.29. Termostat jest regulatorem bezpośredniego działania. Czujnikiem jest w nim mieszek sprężysty 1 wypełniony częściowo cieczą, częściowo parą nasyconą. Zmieniające się, w zależności od temperatury, ciśnienie pary powoduje rozciąganie lub kurczenie mieszka i przemieszczanie połączonego z nim grzybka 2 w gnieździe zaworu 3. W termostacie tym przesunięcie grzybka o ok. 10 mm uzyskuje się przy zmianie temperatury w granicach od ok. 70°C (zawór zamknięty) do ok. 80°C (zawór całkowicie otwarty). A więc termostat jest regulatorem proporcjonalnym mającym ustalony fabrycznie zakres proporcjonalności wynoszący 10°C.

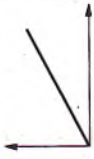
•4.30• Jakimi parametrami charakteryzują się regulatory o działaniu ciągłym?

W tablicy 4.30 zestawiono charakterystyki i podstawowe zależności (transmitancje) i równania charakteryzujące typy



•4.29• Termostat samochodowy

1 — mieszek sprężysty; 2 — grzybek; 3 — gniazdo zaworu

Regulator	Transmitancja operatorowa	Równanie	Charakterystyka skokowa	Charakterystyka amplitudowa logarymiczna (asymptotyczna)	Parametry nastawiane i zakresy zmienności
Proportionalny (typu P)	k_p	$u(t) = k_p \varepsilon$			$x_p = \frac{1}{k_p} 100\%$ 3...500
Całkujący (typu I)	$\frac{1}{T_I s}$	$u(t) = u(0) + \frac{1}{T_I} \int_0^t \varepsilon(\tau) d\tau$			$T_I = 0,1 \dots 30 \text{ min}$
Proportionalno-całkujący (typu PI)	$k_p \left(1 + \frac{1}{T_I s} \right)$	$u(t) = k_p \left(\varepsilon + \frac{1}{T_I} \int_0^t \varepsilon(\tau) d\tau \right) + u(0)$			$x_p = 3 \dots 500$ $T_I = 0,1 \dots 30 \text{ min}$
Proportionalno-całkująco-różniczkujący (typu PID)	$k_p \left(1 + \frac{1}{T_I s} + T_d s \right)$	$u(t) = k_p \left(\varepsilon + \frac{1}{T_I} \int_0^t \varepsilon(\tau) d\tau + T_d \frac{d\varepsilon}{dt} \right) + u(0)$			$x_p = 3 \dots 500$ $T_I = 0,1 \dots 30 \text{ min}$ $T_d = 0 \dots 10 \text{ min}$
Proportionalno-całkująco-różniczkujący (PID) z nieidealnym różniczkowaniem	$k_p \left(1 + \frac{1}{T_I s} + \frac{T_d s}{\tau s + 1} \right)$				$x_p = 3 \dots 500$ $T_I = 0,1 \dots 30 \text{ min}$ $T_d = 0 \dots 10 \text{ min}$

regulatorów stosowanych w przemyśle. Dane te opisują właściwości regulatorów nieco wyidealizowanych. Jednak, mimo iż w wielu przypadkach właściwości rzeczywistych regulatorów różnią się nieco od podanych w tablicy (tak np. właściwości członu całkującego mogą odbiegać od przedstawionych), to przedstawione zależności są wystarczająco dokładne z punktu widzenia praktyki przemysłowej.

Z tablicy 4.30 wynika, że regulatory mają najwyżej trzy dające się nastawiać parametry: **współczynnik wzmocnienia, stałą czasową całkowania i stałą czasową różniczkowania.**

Dla parametrów tych stosuje się następujące nazwy:

k_p — współczynnik wzmocnienia

$\frac{1}{k_p} 100\% = x_p$ — zakres proporcjonalności

T_i — czas zdwojenia

T_d — czas wyprzedzenia

Zakres proporcjonalności podaje, przy jakim zakresie zmian sygnału uchybu (który jest sygnałem wejściowym regulatora) jest zachowana proporcjonalność pomiędzy wartością sygnału uchybu a wartością sygnału wyjściowego regulatora proporcjonalnego.

Czas zdwojenia określa szybkość zmian sygnału wyjściowego regulatora przy stałej wartości sygnału uchybu; jest to czas, po którym sygnał wyjściowy regulatora PI osiąga (po wymuszeniu skokowym) dwukrotnie większą wartość niż miałoby to miejsce w regulatorze typu P. Liczbowo czas zdwojenia jest równy stałej czasowej całkowania członu całkującego. Oczywiście (widać to z zależności podanych w tabl. 4.30) im mniejsza jest wartość czasu zdwojenia, tym intensywniejsze jest całkowanie. **Czas wyprzedzenia** T_d określa właściwości członu różniczkującego regulatora. Dodanie do regulatora członu różniczkującego wprowadza do sygnału wyjściowego składnik zależny od szybkości zmian uchybu. Dzięki temu do układu wprowadza się jak gdyby przewidywanie (wyprzedzenie): jeżeli uchyb narasta i jest coraz większy, regulator wytwarza sygnał przeciwdziałający temu narastaniu wcześniej, niż mogłoby to być zrobione na podstawie znajomości tylko aktualnej wartości uchybu.

4.31. Czy transmitancja regulatora całkowicie charakteryzuje jego właściwości?

Nie. Unowocześnianie regulatorów polega na zwiększaniu dokładności ich pracy, na rozszerzaniu zakresu dodatkowych zadań, jakie mogą one spełniać oraz na udoskonalaniu funkcji dodatkowych już istniejących. A więc dla scharakteryzowania właściwości regulatora należy podać szereg dodatkowych informacji określających:

- sposób i dokładność ustawiania wartości zadanej;
- sposób i dokładność ustawiania nastaw;
- stałość ustawionych wartości nastaw i wartości zadanej, w funkcji czasu i zmian czynników zewnętrznych;

— czy istnieje możliwość przełączania rodzaju pracy regulatora bez wprowadzania zakłóceń do układu regulacji (tzn. czy przy przełączaniu rodzaju pracy regulatora zmienia się jego sygnał wyjściowy);

— czy istnieje możliwość współpracy z maszyną cyfrową;

— ile sygnałów można sumować na wejściu regulatora;

— czy regulator może być wykonany w wersji umożliwiającej pracę w układzie kaskadowym;

— czy istnieje ograniczenie całkowania przy maksymalnej lub minimalnej wartości sygnału regulatora.

Właściwości regulatora, a więc jego przydatność do konkretnego zastosowania, zależy jeszcze od szeregu innych czynników, wśród nich przede wszystkim od ogólnych cech urządzeń systemu blokowego, do którego wchodzi dany regulator.

·4.32· Jak jest zbudowany nowoczesny regulator elektryczny?

Dobrym przykładem elektrycznego regulatora działania ciągłego jest regulator ARC-21 (produkowany przez MERA ELPO, Wrocław). Rysunek 4.23A przedstawia uproszczony schemat tego regulatora w połączeniu ze stacyjką sterowania ręcznego ADS-42. Regulator może także pracować ze stacyjką innego typu — np. ze stacyjką umożliwiającą współpracę z komputerem. Zastosowane do budowy regulatora wzmacniacze scalone i inne elementy układu elektronicznego są wysokiej klasy, odznaczają się dokładnością i stabilnością. Dzięki temu regulator także pracuje dokładnie i niezawodnie.

Transmitancję regulatora można określić na podstawie schematu z rys. 4.32A. Postępując zgodnie z zasadami opisywania transmitancji wzmacniaczy operacyjnych otrzymamy po pewnych uproszczeniach

$$U(s) = k_p \left[1 + \frac{1}{T_i s} \right] \varepsilon(s) - k_p \frac{T_d s}{1 + \frac{T_d}{k_d}} X(s)$$

gdzie:

$$k_p \approx \frac{R_2}{\alpha R_1} \quad \text{— wzmacnienie regulatora;}$$

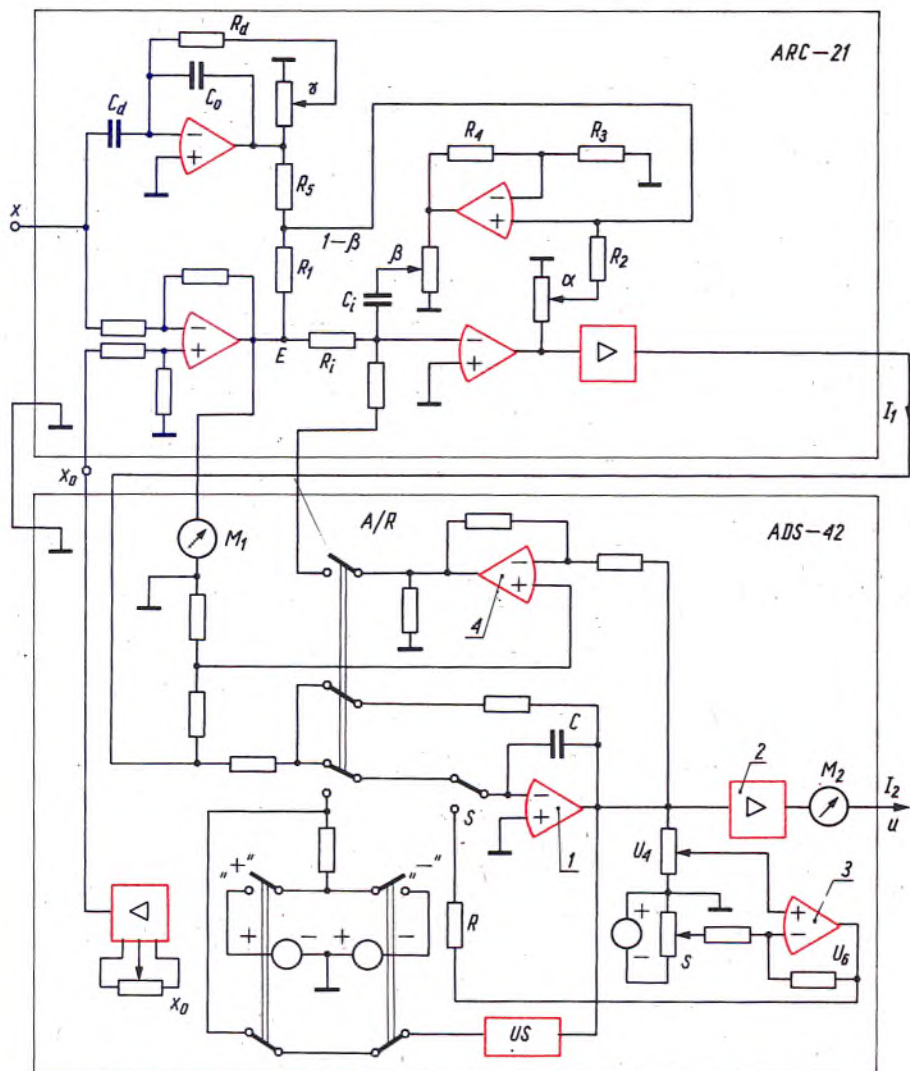
$$T_i = \frac{C_1 R_1 (R_3 + R_4)}{R_3 \left(1 + \frac{R_1}{R_2} + \frac{R_1}{R_2} \right)} \quad \text{— czas zdwojenia regulatora;}$$

$$T_d = \frac{R_d C_d R_1}{\gamma R_5} \quad \text{— czas wyprzedzenia regulatora;}$$

$$k_d = \frac{C_d R_1}{C_0 R_5} \approx 17; \quad \varepsilon(s) = X_0(s) - X(s) \quad \text{— transformata uchybu regulacji}$$

$U(s)$ — transformata sygnału wyjściowego regulatora.

Jest to więc regulator PID z modyfikacją: różniczkowaniu podlega nie uchyb regulacji (jak w większości regulatorów),



•4.32A• Uproszczony schemat regulatora ARC-21 w połączeniu ze stacyjką sterowania ręcznego ADS-42

1 — wzmacniacz pracujący w układzie integratora przy sterowaniu ręcznym; 2 — wyjściowy wzmacniacz prądowy; 3 — wzmacniacz pracujący w układzie ustawiania wartości „S”; 4 — wzmacniacz pracujący w układzie sprzężenia pomiędzy członem całkującym a sygnałem sterowania ręcznego



•4.32B• Stacyjka sterowania ręcznego ADS-42

a sygnał wyjściowy obiektu. Jeżeli wartość zadana jest stała, to oczywiście jest wszystko jedno, który z tych dwu sygnałów jest różniczkowany; natomiast przy szybkich zmianach wartości zadanej rozwiązanie przyjęte w regulatorze ARC-21 jest korzystniejsze, ponieważ zapobiega wchodzeniu regulatora w stan nasycenia przy zmianach wartości zadanej. Nastawy regulatora ARC-21 można zmieniać w szerokich granicach, przyjętych dla regulatorów dobrej klasy: k_p od 0,3 do 300, T_i od 0,1 do 30 min., T_d od 0 do 10 min. Obwody wejściowe i wyjściowe mogą być przystosowane do pracy z sygnałem prądowym o zakresie zmienności od 0 do 5 mA lub od 4 do 20 mA. Układem wyjściowym regulatora jest wzmacniacz prądowy (umieszczony w stacyjce sterowania ręcznego), dzięki temu prąd wyjściowy nie zależy od wartości rezystancji obciążenia, jeżeli jest ona mniejsza niż $2\text{ k}\Omega$.

Potencjometry i przełączniki umożliwiające zmiany nastaw regulatora ARC 21 są umieszczone w regulatorze, wszystkie manipulacje pomocnicze wykonuje się za pomocą elementów stanowiących wyposażenie stacyjki. Stacyjka sterowania ręcznego ADS 42 (rys. 4.32B) spełnia szereg ważnych funkcji. Zawiera ona:

- Nadajnik wartości zadanej x_0 . Jest nim precyzyjny potencjometr dziesięcioobrotowy, współpracujący ze stabilizowanym źródłem prądu stałego.

- Przełącznik praca automatyczna-sterowanie ręczne (przełącznik A/R). Przy pracy automatycznej (pozycja przełącznika jak na rys. 4.32A) wzmacniacz pracuje jak człon inercyjny o bardzo małej stałej czasowej. Prąd wyjściowy stacyjki, będący sygnałem wyjściowym regulatora, jest wtedy równy prądowi wyjściowemu regulatora I_1 . Przy sterowaniu ręcznym wzmacniacz pracuje jak integrator z układem stabilizacji (blok US). Jednocześnie wartość sygnału wyjściowego regulatora jest podawana do członu całkującego.

- Przyciski sterowania ręcznego. Naciśnięcie przycisków „+” lub „-” powoduje przyłączanie do wejścia wzmacniacza I (który przy pracy ręcznej jest integratorem) stałego napięcia. Na skutek tego sygnał wyjściowy integratora, a więc również sygnał wyjściowy regulatora, zmienia się liniowo. Zmiana sygnału wyjściowego od zera do wartości maksymalnej trwa 60 s. Naciskanie przycisków sterowania ręcznego powoduje także odłączanie układu stabilizacji napięcia wyjściowego integratora.

- Przełącznik S . Wciśnięcie tego przełącznika powoduje, że sygnał wyjściowy regulatora zaczyna zmieniać się liniowo, przyjmując ostatecznie wartość ustawioną wcześniej za pomocą dziesięcioobrotowego precyzyjnego potencjometru „S”. Ten rodzaj pracy regulatora jest pożyteczny wówczas, gdy w przypadku uszkodzenia układu pomiarowego lub regulatora trzeba szybko ustawić element wykonawczy w wybranej wcześniej pozycji „bezpiecznej”.

- Miernik uchybu M_1 .

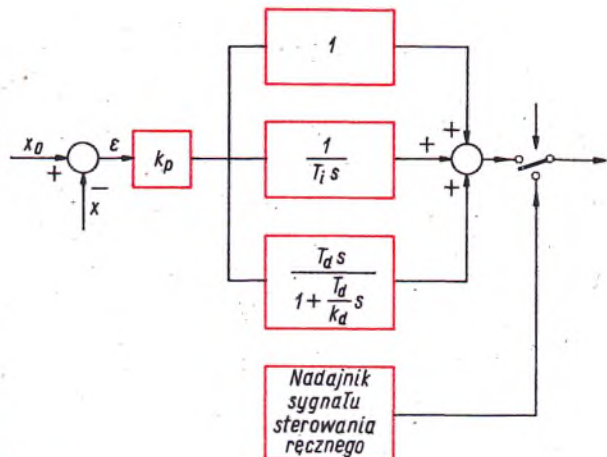
- Miernik prądu wyjściowego M_2 .

4.33. W jaki sposób uzyskuje się przełączanie rodzaju pracy regulatora bez wprowadzania zakłóceń do układu regulacji?

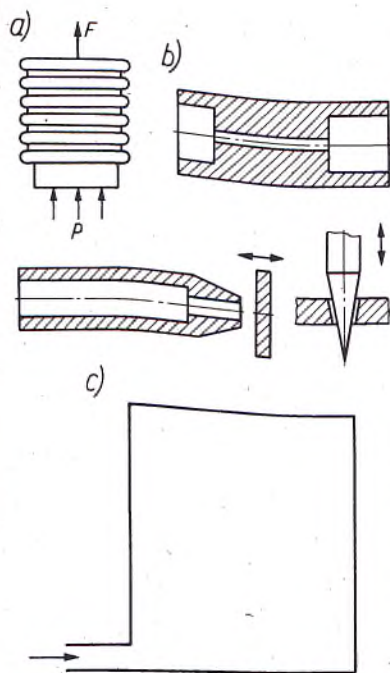
Tak zwane **bezzakłócenkowe przełączanie** rodzaju pracy regulatora uzyskuje się pod warunkiem niezmienności jego sygnału wyjściowego w chwili przełączania.

Bezzakłócenkowe przełączanie z pracy automatycznej na sterowanie ręczne jest zapewnione, jeżeli sterowanie ręczne polega na zmienianiu, a nie na ustawianiu wartości sygnału wyjściowego. Tak właśnie jest w regulatorze ARC-21 (rys. 4.32A). Po przełączeniu na sterowanie ręczne, w integratorze zawierającym wzmacniacz I jest zapamiętana wartość sygnału wyjściowego, istniejąca w chwili przełączania. Wartość ta jest utrzymywana i dopiero naciśnięcie jednego z przycisków sterowania ręcznego może ją zmienić. Gdy wartość sygnału sterowania ręcznego jest ustawiana, np. za pomocą potencjometru, wówczas przed przełączeniem na sterowanie ręczne należy ją zrównać z wartością sygnału wyjściowego regulatora.

Przełączenie na pracę automatyczną rozpatrzmy na schemacie blokowym regulatora PID, przedstawionym na rys. 4.33. Sygnał wyjściowy regulatora jest sumą sygnałów członów: proporcjonalnego, różniczkującego i całkującego. Jeżeli uchyb regulacji jest równy zeru, to sygnały wyjściowe członu proporcjonalnego i różniczkującego także są równe zeru. Sygnał wyjściowy regulatora jest więc wówczas równy sygnałowi wyjściowemu członu całkującego. Aby uzyskać bezzakłócenkowe przełączanie ze sterowania ręcznego na pracę automatyczną, należy zapewnić, aby sygnał wyjściowy członu całkującego był w chwili przełączania równy sygnałowi wyjściowemu regulatora. W regulatorze ARC 21 przy sterowaniu ręcznym istnieje sprzężenie zwrotne (poprzez wzmacniacz 4) pomiędzy wyjściem regulatora a członem całkującym. Dzięki temu sygnał wyjściowy członu całkującego przy ste-



4.33. Schemat blokowy regulatora PID



·4.34· Elementy aparatury pneumatycznej: a) mieszek sprężysty; b) opory pneumatyczne; c) komora

rowaniu ręcznym jest stale równy sygnałowi wyjściowemu regulatora.

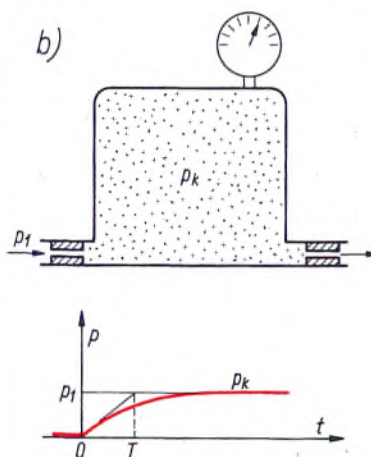
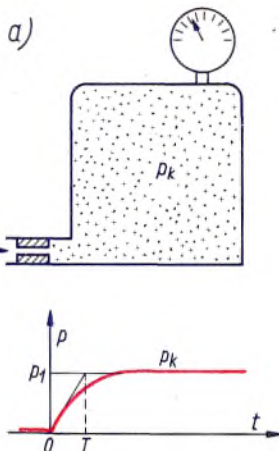
·4.34· Jakie są najważniejsze elementy pneumatyczne potrzebne do budowy urządzeń automatyki?

W pneumatycznych urządzeniach automatyki przeważają systemy siłowe, w których porównywanie i sumowanie sygnałów pneumatycznych odbywa się po uprzednim przetworzeniu ich na sygnały siłowe. Do zmiany ciśnień na siły służą **membrany** oraz **mieszki sprężyste**; większe zastosowanie mają mieszki.

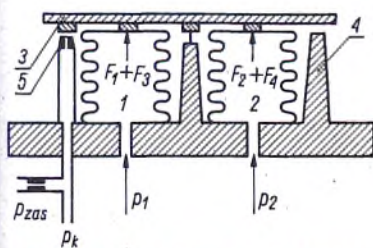
Mieszki (rys. 4.34) są wykonywane z mosiądzu lub brązu berylowego, ich właściwości charakteryzuje powierzchnia czynna i sztywność. Powierzchnia czynna mieszka jest w przybliżeniu równa powierzchni denka i nie zależy od zmiany jego długości. Sztywność (sprężystość) mieszka zależy od właściwości materiału i można ją określić tylko doświadczalnie. Dla zastosowań w urządzeniach automatyki bardzo ważna jest niezmienność właściwości mieszka po jego zamontowaniu w urządzeniu. W urządzeniach siłowych zmiany długości mieszków są ograniczone do wartości rzędu dziesiętnych części milimetra. Przy tak małych odkształceniach zmiany sprężystości mieszka są pomijalnie małe, a ponadto mieszki są zabezpieczone przed trwałymi odkształceniami. Siła wywierana przez mieszek jest równa iloczynowi powierzchni czynnej mieszka i różnicy ciśnień panujących wewnątrz i na zewnątrz mieszka. Bardzo ważną rolę odgrywają w urządzeniach pneumatycznych stale i zmienne **opory pneumatyczne** (rys. 4.34). Są to elementy ograniczające przepływ powietrza i w związku z tym powodujące spadki ciśnienia. Realizuje się je jako zwężenia przewodów — wartość oporu zależy od długości i przekroju tego zwężenia. W oporach stałych długość i przekrój zwężenia są wielkościami stałymi, w oporach zmiennych jedna lub obie te wielkości są zmieniane. Zmiany wartości oporów przedstawionych na rys. 4.34 powodowane są zmianami parametrów zwężenia wywołanymi przesuwaniem elementów ruchomych, tak jak wskazują strzałki na rysunku. Opór pneumatyczny składający się z dyszy i przysłony jest stosowany powszechnie tam, gdzie istnieje potrzeba zamiany przesunięcia na wartość ciśnienia. Wartość tego oporu zależy od położenia przysłony względem dyszy. Zespół dysza-przysłona jest używany zwykle jako pneumatyczny opór zmienny tłumiący wpływ powietrza do atmosfery.

Komory (zbiorniki) pneumatyczne umożliwiają, w połączeniu z oporami pneumatycznymi, tworzenia członów dynamicznych. Spełniają one w układach pneumatycznych tę samą rolę, co pojemności w układach elektrycznych.

Komory mogą być **sztywne** (o stałej objętości) bądź **elastyczne** (o objętości zależnej od ciśnienia). W układach siłowych mieszki są uważane za komory sztywne.



•4.35• Pneumatyczne człony inercyjne oraz ich charakterystyki skokowe: a) z komorą zamkniętą; b) z komorą przelotową



•4.36A• Pneumatyczny zespół porównujący w wykonaniu mieszkowym

1, 2 — mieszki; 3 — przysłona; 4 — zderzak; 5 — dysza

•4.35• Jak tworzy się pneumatyczne człony dynamiczne?

Doprowadzając powietrze do komory pneumatycznej przez opór (rys. 4.35a) otrzymamy człon, którego charakterystykę skokową przedstawia rys. 4.35b. Charakterystyka ta pokazuje zmiany ciśnienia w komorze po skokowej zmianie ciśnienia wejściowego. Jest to charakterystyka członu inercyjnego, a więc łącząc szeregowo opór i pojemność pneumatyczną, otrzymujemy pneumatyczny człon inercyjny. Podobny rezultat otrzymamy przyłączając komorę pomiędzy dwa opory pneumatyczne (rys. 4.25b). Tego typu człony stosowane jako elementy dodatniego lub ujemnego sprzężenia zwrotnego, obejmującego wzmacniacz pneumatyczny, umożliwiają uzyskiwanie w regulatorach pneumatycznych działania całkującego lub różniczkującego.

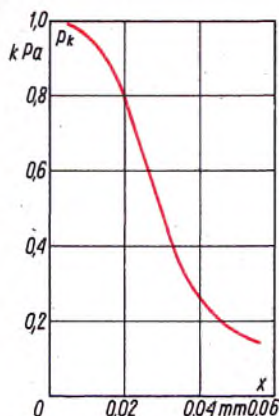
•4.36• Jak realizuje się porównywanie i sumowanie sygnałów w urządzeniach pneumatycznych?

Porównywanie oraz sumowanie sygnałów jest jedną z często wykonywanych operacji. Zasadę działania pneumatycznego członu porównującego typu siłowego rozpatrzmy dokładnie na przykładzie zespołu mieszkowego tzw. równoważni pneumatycznej przedstawionej na rys. 4.36A. Przedstawiona wersja zawiera dwa mieszki ale może być ich więcej.

Wierchołki mieszków 1 i 2 podpierają po przeciwnych stronach osi obrotu dźwigni 3. Wychylenie dźwigni jest ograniczone z jednej strony zderzakiem 4, a z drugiej wylotem dyszy 5. Na dźwigni działają siły F_1 i F_2 , wywołane sygnałami ciśnieniowymi p_1 i p_2 , oraz siły sprężystości mieszków.

Dźwignia znajduje się w równowadze, jeżeli momenty działające po obu stronach jej osi obrotu są sobie równe. Zwykle mieszki są instalowane parami w jednakowych odległościach od osi obrotu tak, że ich siły sprężystości znoszą się wzajemnie, a zmiany tych sił można pominąć. Ponieważ mieszki są jednakowe, więc równość momentów nastąpi przy równości ciśnień doprowadzonych do obu mieszków. Jeżeli równoważnia zawiera więcej mieszków, to należy uwzględnić momenty wprowadzane przez wszystkie mieszki.

Jeżeli momenty działające z obu stron nie są sobie równe, wówczas powstaje moment wypadkowy powodujący odchylenie dźwigni od położenia środkowego. Na dźwigni jest umieszczona przysłona współpracująca z dyszą 5. Zmiany położenia dźwigni zmieniają jej odległość od dyszy, a więc opór pneumatyczny układu dysza-przysłona. Wielkością wyjściową równoważni jest ciśnienie p_k , jego wartość wynika z podziału ciśnienia zasilania w stosunku równym stosunkowi oporów pneumatycznych: wejściowego i zespołu dysza-przysłona. Zależność ciśnienia p_k od położenia przysłony przedstawiono na rys. 4.36B. Zakres zmian odległości przysłony potrzebnych do uzyskania określonych zmian ciśnienia kaskadowego zależy od wymiaru dysz. Dla powszechnie



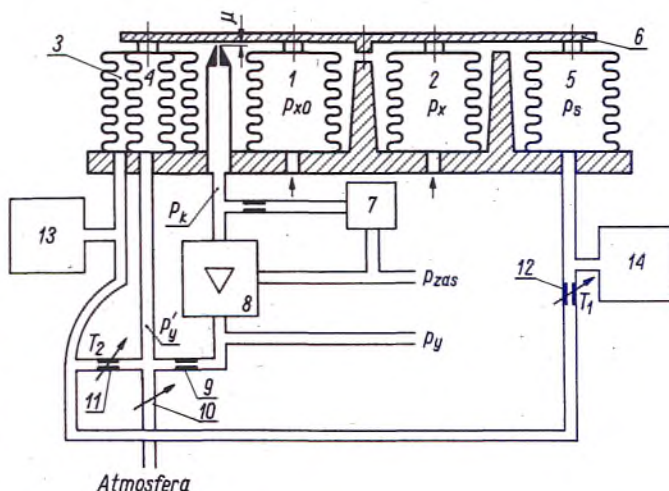
·4.36B· Zależność ciśnienia wyjściowego od odległości dysza-przysłona przy średnicy dyszy 0,05 mm i ciśnieniu zasilania 1 kPa

stosowanych dysz o średnicach wewnętrznych 0,3... 1 mm, zmiany ciśnienia kaskadowego w zakresie 20 ... 80% p_{zas} uzyskuje się przy przesunięciach przysłony rzędu kilku setnych części milimetra. Ponieważ z reguły korzysta się tylko z części charakterystyki układu dysza-przysłona, więc w rzeczywistości przesunięcia są jeszcze mniejsze. Dlatego właśnie można mówić o systemie siłowym i pomijać zmiany naprężeń mieszków.

·4.37· Jak działa regulator pneumatyczny?

Jako przykład regulatora pneumatycznego omówimy mieszkowy regulator PID systemu Telepneu (produkcji MERA-PAP, Falenica). Schemat regulatora przedstawiono na rys. 4.37A. Ciśnienie p_{x0} , reprezentujące wartość zadaną regulatora, oraz ciśnienie p_x reprezentujące wartość mierzoną, oddziałują na mieszki wejściowe 1 i 2. Różnica ciśnień $p_{x0} - p_x$ wyrażająca uchyb regulacji powoduje odchylenie dźwigni 6, co pociąga za sobą zmianę ciśnienia kaskadowego p_k . W regulatorze typu PID ciśnienie zasilające dyszę podlega stabilizacji w stabilizatorze pomocniczym 7. Zmniejsza to wrażliwość regulatora na wahania ciśnienia zasilającego. Ciśnienie kaskadowe p_k steruje wzmacniaczem 8, którego ciśnienie wyjściowe jest ciśnieniem wyjściowym regulatora. Ciśnienie wyjściowe p_y jest dzielone za pomocą dławików 9 i 10, przy czym dławik 9 jest stały, a dławik 10 jest wykonany w formie zaworu igłowego. Dławik 10 służy do ustawiania w regulatorze żądanej wartości wzmocnienia k_p . Otrzymane z dzielnika (zbudowanego z dławików 9 i 10) ciśnienie p_y podawane jest do mieszka wewnętrznego 4, stanowiącego w tym regulatorze mieszek ujemnego sprzężenia zwrotnego.

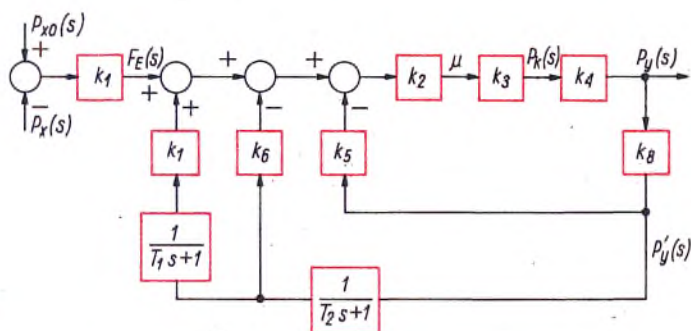
Ciśnienie p_y jest doprowadzane także do mieszków 3 i 5. Do mieszka 3 przez dławik nastawny 11 i komorę 13, a do mieszka 5 przez dławiki nastawne 11 i 12 oraz przez komorę



·4.37A· Regulator mieszkowy typu PID

1, 2 — mieszki wejściowe; 3 — mieszek ujemnego sprzężenia zwrotnego; 4 — mieszek wewnętrzny; 5 — mieszek dodatniego sprzężenia zwrotnego; 6 — dźwignia; 7 — stabilizator ciśnienia; 8 — wzmacniacz; 9 — dławik stały; 10, 11, 12 — dławiki zmienne; 13, 14 — komory

•4.37B• Schemat blokowy regulatora pneumatycznego PID



14. Mamy tu więc do czynienia z ujemnym (mieszek 3) i dodatnim (mieszek 5) sprzężeniem zwrotnym realizowanym poprzez człony inercyjne.

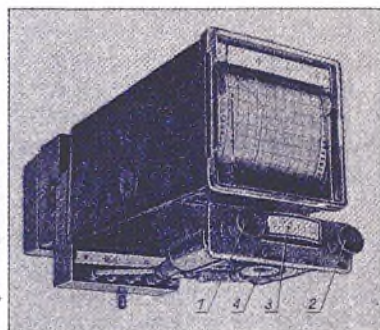
Na podstawie przedstawionego schematu budowy regulatora można narysować jego schemat blokowy (rys. 4.37B) umożliwiający znalezienie transmitancji regulatora. Po przekształceniach i pewnych uproszczeniach otrzymujemy

$$U(s) = \frac{k_1(T_1 + T_2)}{k_7 k_8 T_1(1 + T_3 s)} \left[1 + \frac{1}{(T_1 + T_2)s} + \right. \\ \left. + \frac{T_1 T_2}{(T_1 + T_2)s} \right] \varepsilon(s)$$

przy czym: $k_7 = k_5 + k_6$, $T_3 = \frac{k_5 T_1 T_2}{T_1 k_7 + T_2 k_5}$; pozostałe oznaczenia jak na schemacie blokowym.

Przedstawione zależności pokazują, że w regulatorze występuje **interakcja** czyli wzajemna zależność nastaw. Istotnie, zmieniając za pomocą dławika II wartość T_2 nastawiamy wartość czasu wyprzedzenia ale jednocześnie zmienia się wartość czasu zdwojenia. Ograniczeniem wynikającym z interakcji jest m.in. niemożność ustawienia regulatora na działanie typu PD. Działanie typu PI teoretycznie można uzyskać otwierając całkowicie dławik II.

Pod względem konstrukcji regulator jest wykonany z przeznaczeniem do mocowania na korpusie rejestratora lub na korpusie miernika wskazówkowego. Na rysunku 4.37C przedstawiono regulator zamocowany na korpusie rejestratora, połączony ze stacyjką sterowania ręcznego. Zastosowana tu stacyjka sterowania ręcznego najprostszego typu jest wyposażona w nastawnik wartości zadanej 1, nastawnik sterowania ręcznego 2, przyrząd (manometr) dwuwskazówkowy 3 i przełącznik rodzaju pracy 4. Przyrząd dwuwskazówkowy umieszczony w stacyjce wskazuje zwykle wartość ciśnienia wyjściowego regulatora oraz wartość ciśnienia wyjściowego nastawnika sterowania ręcznego. Przy przełączaniu ze sterowania automatycznego na sterowanie ręczne obie wskazówki powinny wskazywać tę samą wartość, przełączenie nie powoduje wtedy zakłóceń w pracy obiektu.



•4.37C• Regulator pneumatyczny zamontowany na korpusie rejestratora

1 — nastawnik wartości zadanej; 2 — nastawnik sterowania ręcznego; 3 — przyrząd dwuwskazówkowy; 4 — przełącznik rodzaju pracy

•4.38• Jak są budowane regulatory hydrauliczne?

Regulatory hydrauliczne nie istnieją jako specjalne urządzenia ogólnego przeznaczenia. Urządzenia nazywane regulatorami hydraulicznymi powstają przez połączenie wzmacniacza hydraulicznego z siłownikiem hydraulicznym i doprowadzenie do wejścia sterującego wzmacniacza sygnału wartości wielkości mierzonej. Przez proste połączenie wzmacniacza z siłownikiem otrzymujemy regulator typu I. Stosując sprzężenie zwrotne obejmujące siłownik można używać regulator typu P (jeżeli zastosujemy sprzężenie sztywne), lub regulator typu PI — jeżeli sprzężenie zwrotne zostanie zrealizowane z użyciem członu zwanego izodromem. Ze względu na duże trudności w realizacji hydraulicznych członów czasowych regulatory bardziej skomplikowane, niż typu PI nie są w zasadzie stosowane.

Regulatory hydrauliczne buduje się w postaci bloków, w których wzmacniacz, siłownik i elementy sprzężenia zwrotnego tworzą konstrukcyjnie całość. Taki zespół regulacyjny jest wyposażony często we własny zespół zasilający go cieczą o wysokim ciśnieniu.

Regulatory hydrauliczne ze względu na małą uniwersalność i wysoką cenę są coraz rzadziej używane.

•4.39• Jak działa układ regulacji z regulatorem hydraulicznym?

Jako przykład zastosowania regulatora hydraulicznego, na rys. 4.39 przedstawiono schemat układu regulacji stosunku przepływów dwu gazów. Może to być np. układ regulacji stosunku dopływu gazu i powietrza do pieca opalanego gazem.

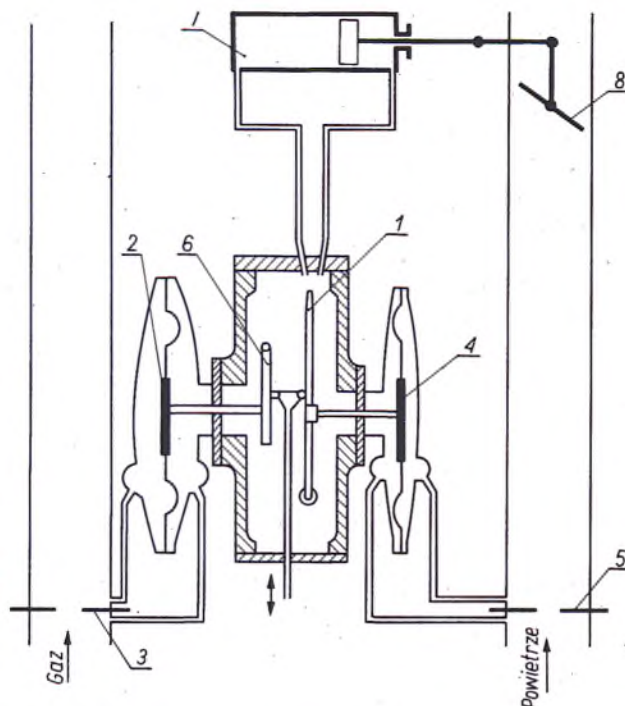
Przedstawiony układ stanowi w istocie regulator przepływu powietrza, w którym wartością zadaną jest natężenie przepływu gazu. Regulator zbudowano przy użyciu wzmacniacza z rurką strumieniową. Na rurkę 1 działa z jednej strony siła nacisku membrany pomiarowej 2 z układu pomiaru różnicy ciśnień na zwężce pomiarowej natężenia przepływu gazu 3. Z drugiej strony działa siła takiej samej membrany pomiarowej 4 pracującej w układzie pomiaru natężenia przepływu powietrza. Siły obu membran działają przez układ dźwigniowy, którego przełożenie można zmieniać zmieniając położenie ślizgacza 6. Ciśnienie wyjściowe rurki strumieniowej steruje siłownikiem 7 przedstawiającym przepustnicę 8.

W układzie regulatora przepływu powietrza nie ma wartości zadanej natężenia przepływu powietrza, nie jest to więc układ regulacji stałowartościowej. Jest to układ **regulacji nadążnej**, a wielkość zastępującą wartość zadaną nazywamy wielkością **wiodącą**. W naszym przykładzie wielkością wiodącą jest natężenie przepływu gazu.

Wartość zadana stosunku przepływów gazu i powietrza jest określona położeniem ślizgacza 6. Od położenia tego

•4.39• Hydrauliczny układ regulacji stosunku przepływów dwu gazów

1 — rurka strumieniowa; 2 — membrana pomiarowa; 3 — zwężka pomiarowa; 4 — membrana pomiarowa; 5 — zwężka pomiarowa; 6 — ślizgacz ustawnika wartości stosunku; 7 — siłownik; 8 — przepustnica



ślizgacza zależy przełożenie układu dźwigniowego dla sił wywieranych przez membrany 2 i 4 — wraz ze zmianami położenia ślizgacza 6 zmieniają się bowiem długości ramion, na jakich działają siły od obu membran. Zmieniając położenie ślizgacza można zmieniać (w zakresie od 0,25 do 4) stosunek sił od obu membran, przy którym rurka strumieniowa znajduje się w środkowym położeniu. Daje to możliwość ustawiania wartości zadanej stosunku natężeń w granicach od 0,5 do 2.

Przedstawiony regulator stosunku jest regulatorem typu I. Jeżeli rurka strumieniowa jest choćby minimalnie odchylona od położenia środkowego, występuje różnica ciśnień po obu stronach tłoka siłownika i tłok wykonuje ruch przestawiający przepustnicę 8. Warunkiem równowagi układu jest zgodność stosunku przepływów obu czynników z wartością zadaną — a więc zerowy uchyb regulacji.

•4.40• W jakim celu w układach regulacji automatycznej stosuje się rejestratory?

Rejestracja wartości ważnych wielkości charakteryzujących proces technologiczny ułatwia pracę operatorowi nadzorującemu przebieg tego procesu, ponieważ umożliwia śledzenie zmian tych wielkości w funkcji czasu. Ponadto zapis z rejestratora służy często jako konieczna dokumentacja przebiegu procesu. Te same funkcje spełniają rejestratory w układach regulacji automatycznej. Niezależnie bowiem od stosowania układów regulacji automatycznej, procesy

technologiczne są prowadzone z reguły pod nadzorem operatorów. Klasyczne środki automatyki nie zapewniają bowiem właściwego przebiegu procesów w sytuacjach wyjątkowych. Takimi sytuacjami wyjątkowymi mogą być duże zakłócenia, uszkodzenia przyrządów itp. Krótko — rejestratory przemysłowe służą do zapisywania zmian wielkości mierzonej w czasie.

·4.41· Jakiego rodzaju rejestratory są stosowane w układach regulacji automatycznej

Ponieważ zadaniem rejestratorów w układach automatyki jest rejestracja wartości wielkości mierzonej w funkcji czasu, więc każdy rejestrator zawiera przynajmniej dwa mechanizmy: mechanizm pomiarowy i mechanizm przesuwu papieru. Buduje się też rejestratory do rejestracji dwóch, a nawet trzech wielkości i zawierające w związku z tym dwa lub trzy mechanizmy pomiarowe.

Ze względu na budowę mechanizmu pomiarowego stosowane obecnie w przemyśle rejestratory można podzielić na — Rejestratory pneumatyczne bezpośredniego działania (tzn. takie, w których element pomiarowy bezpośrednic napędza pisak);

— Rejestratory elektryczne bezpośredniego działania z zapisem punktowym;

— Rejestratory elektryczne z zapisem ciągłym, autokompensacyjne.

Przesuw papieru odbywa się zawsze ze stałą prędkością; podstawowym warunkiem stawianym mechanizmom przesuwu jest zachowanie równomierności przesuwu. Stosowane są dwa rodzaje mechanizmów przesuwu: do zapisu na przesuwanej taśmie lub na obracanej tarczy papierowej.

·4.42· Jakiego rodzaju nowoczesne metody rejestracji stosuje się w układach automatyki?

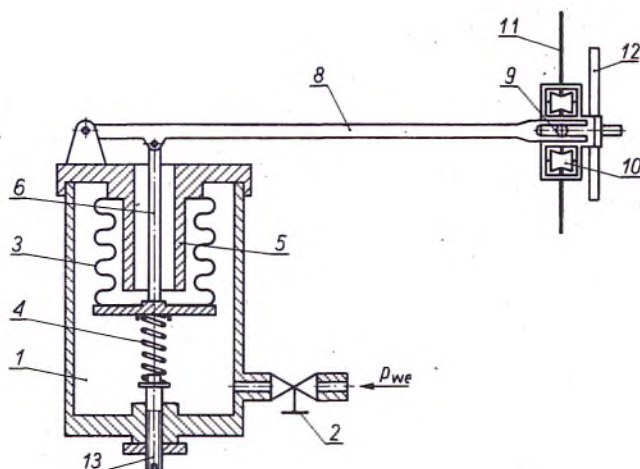
Są to metody cyfrowe. Przy aktualnym poziomie techniki cyfrowej można stosować układy cyfrowe również w przemyśle. Układy cyfrowe, znane pod nazwą układów **centralnej rejestracji i przetwarzania danych** (w skrócie CRPD), umożliwiają dokładny pomiar i wygodne dla operatora sposoby rejestracji wielu wielkości. Dokładniejszych informacji na ten temat należy szukać w rozdz. 5 poświęconym urządzeniom cyfrowym.

·4.43· Jak jest zbudowany rejestrator pneumatyczny?

W rejestratorach pneumatycznych najczęściej stosuje się do pomiaru ciśnienia układy mieszkowe (rys. 4.43). Rejestrowany sygnał ciśnieniowy jest doprowadzany do komory 1 przez zawór dławiący 2, zawór ten wraz z pojemnością komory tworzy człon inercyjny tłumiący tętnienia sygnału rejestrowanego. Ciśnienie panujące w komorze działa na mieszek sprężysty 3, siła pochodząca od ciśnienia równoważona jest

4.43. Mechanizm pomiarowy rejestratora pneumatycznego

1 — komora; 2 — zawór dławiący; 3 — mieszek; 4 — sprężyna pomiarowa; 5 — ograniczenie ruchu mieszka; 6 — popychacz; 7 — dźwignia; 8 — wskazówka; 9 — widelki; 10 — karetki; 11 — strona jezdna; 12 — prowadnica; 13 — śruba regulacyjna



przez sprężynę pomiarową 4. Ukształtowanie denka komory w formie rurki 5 ogranicza odkształcenie mieszka i zabezpiecza go w ten sposób przed przeciążeniem. Przytwierdzony do denka mieszka popychacz 6 działa na dźwignię 7, prostoliniowe przemieszczenia popychacza są zamieniane przy tym na ruch obrotowy dźwigni. Z dźwigni 7 jest połączona tylna część wskazówki 8, połączenie spełnia rolę bezpiecznika przeciążeniowego. Wskazówka jest zakończona widelkami 9 obejmującymi bolec karetki 10. Karetki porusza się po stronie jezdnej 11, wykonanej z naciągniętego drutu, a jest podtrzymywana przez prowadnicę 12. Na karetkce jest zawieszony pisak (nie pokazany na rysunku), połączony giętą rurką włoskową z zbiornikiem atramentu.

Doprowadzenie sygnału wejściowego do komory powoduje ugięcie mieszka, proporcjonalne do wartości sygnału, obrót dźwigni 7 i połączonej z nią wskazówki 8, a w konsekwencji przesunięcia karetki wzdłuż prowadnicy. Zakres pomiarowy rejestratora można regulować przez zmianę naciągu sprężyny 4 za pomocą śruby regulacyjnej 13.

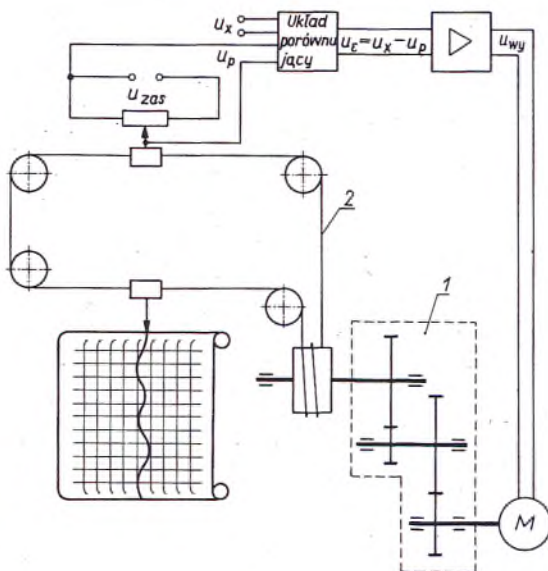
4.44. Jak działa elektryczny rejestrator autokompensacyjny?

Rejestratory autokompensacyjne (zwane autokompensatorami) zapewniają większą dokładność niż inne typy rejestratorów dzięki temu, że do przesuwania pisaków stosuje się w nich dokładne serwomechanizmy. Zasadę działania autokompensacyjnego rejestratora sygnałów napięciowych wyjaśnia rys. 4.44A.

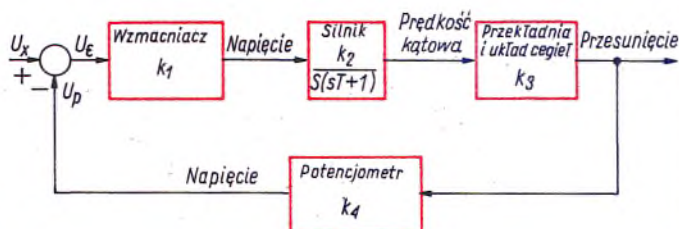
Sygnał wejściowy u_x jest podawany na wejście członu porównującego i tam jest porównywany z napięciem u_p otrzymanym z suwaka potencjometru. Jeżeli potencjometr jest wykonany starannie, a człon porównujący ma dużą rezystancję wejściową, można uzyskać proporcjonalność pomiędzy przesunięciem suwaka potencjometru a napięciem u_p .

•4.44A• Schemat działania rejestratora autokompensacyjnego

1 — przekładnia zębata; 2 — cięgło linkowe;
 M — serwowmotor



Na wyjściu członu porównującego otrzymuje się napięcie $u_e = u_x - u_p$, które jest wzmacniane we wzmacniaczu mocy. Napięcie wyjściowe wzmacniacza zasila silnik M zwany serwowmotorem. Silnik przez przekładnię 1 i cięgło linkowe 2 przesuwa suwak potencjometru w takim kierunku, że maleje różnica między napięciem u_x i u_p . Silnik zatrzymuje się dopiero wtedy, gdy napięcie u_e jest równe zeru, czyli gdy zostanie uzyskana odpowiedniość między sygnałem wejściowym u_x a położeniem suwaka potencjometru. Cięgło linkowe 2 przesuując suwak potencjometru przesuwa także pisak. Jeżeli sygnał wyjściowy u_x zmienia się, to suwak potencjometru jest przesuwany tak, aby zmiany napięcia odniesienia u_p nadążały za zmianami sygnału wejściowego. Przesuwany wraz z suwakiem pisak rejestratora zapisuje, na przewijanym ze stałą prędkością papierze, wykres zmian sygnału wejściowego w funkcji czasu. Dokładność serwomechanizmu przesuującego pisak rejestratora, a więc dokładność pomiaru przesunięcia oraz od występujących w układzie napędowym luzów, tarcia itp. Schemat blokowy rejestratora przedstawiono na rys. 4.44B. Wielkością regulowaną jest w tym układzie położenie pisaka. Wielkością wejściową jest sygnał rejestrowany, mający po-



•4.44B• Schemat blokowy rejestratora autokompensacyjnego

stać napięcia. Zwróćmy uwagę, że sygnał wejściowy z założenia ulega ciągłym zmianom, a zadaniem układu rejestratora jest nadążanie (śledzenie) za tymi zmianami. Napięcie u_p jest miarą wielkości regulowanej, a otrzymywany z członu porównującego sygnał u_e jest sygnałem uchybu.

W stanie ustalonym wartość uchybu powinna być równa zeru tzn., że przy stałej wartości sygnału rejestrowanego pisak powinien zajmować położenie dokładnie odpowiadające tej wartości. Zerowa wartość uchybu w stanie ustalonym wynika wprost z astatyczności układu czyli z istnienia w układzie członu całkującego. W praktyce dokładność rejestracji jest ograniczona dokładnością potencjometru pomiarowego, luzami w mechanizmach, dokładnością przesuwu papieru itp. Ponadto na dokładność rejestracji mają wpływ błędy dynamiczne. Dla omawianego rejestratora wartość uchybu dynamicznego jest proporcjonalna do szybkości zmian sygnału rejestrowanego. W przemysłowych zastosowaniach rejestratorów uchyby dynamiczne na ogół można pominąć, ponieważ sygnały podlegające rejestracji zmieniają się bardzo wolno.

•4.45• Według jakich reguł dobiera się nastawy regulatorów?

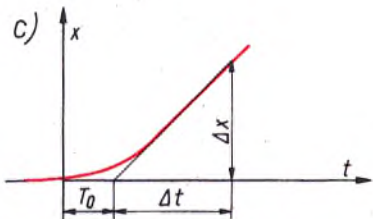
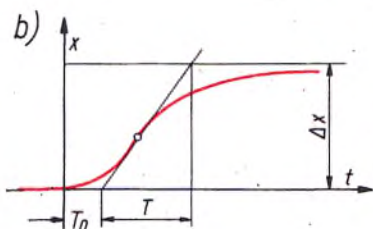
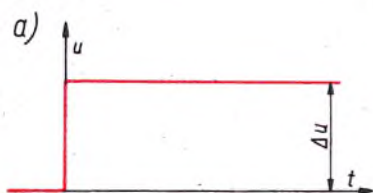
Metody ścisłej analizy teoretycznej, które umożliwiają wyznaczenie dowolnych właściwości projektowanego układu regulacji, nie są zbyt przydatne do dobierania wartości parametrów (tzn. wartości nastaw) regulatorów uniwersalnych. Rzadko bowiem, mając zadanie ustalenia wartości nastaw regulatora, dysponujemy jednocześnie dokładnymi danymi obiektu i założeniami sprecyzowanymi w sensie matematycznym. Najczęstszym zadaniem jest ustalenie takich nastaw regulatora, przy których uzyskuje się zadowalające (w potocznym sensie tego słowa) działanie układu.

Najbardziej rozpowszechnionymi metodami doboru nastaw są: metoda Zieglera-Nicholsa oraz metody oparte na pomiarze charakterystyk skokowych.

Stosując metodę Zieglera-Nicholsa, regulator połączony z obiektem ustawia się na działanie typu P. Następnie zwiększając stopniowo wzmocnienie tego regulatora doprowadza się układ do granicy stabilności. Notując wartość współczynnika wzmocnienia regulatora k_{kr} , przy którym powstają oscylacje, oraz okres t_{osc} tych oscylacji, przyjmuje się odpowiednio

$$\begin{aligned} \text{dla regulatora typu P} & \quad - k_p = 0,5k_{kr} \\ \text{dla regulatora typu PI} & \quad - k_p = 0,45k_{kr}, \quad T_i = 0,85t_{osc}, \\ \text{dla regulatora typu PID} & \quad - k_p = 0,6k_{kr}, \quad T_i = 0,5t_{osc}, \\ & \quad \quad \quad T_d = 0,12t_{osc} \end{aligned}$$

Metoda Zieglera-Nicholsa ma ścisły związek z korekcją charakterystyk dynamicznych (por. pyt. 3.37 ... 3.30). Uzasadnia ona zalecaną wartość wzmocnienia regulatora od wzmocnienia krytycznego układu bez korygowanych charakterystyk dynamicznych oraz punkty załamania charakte-



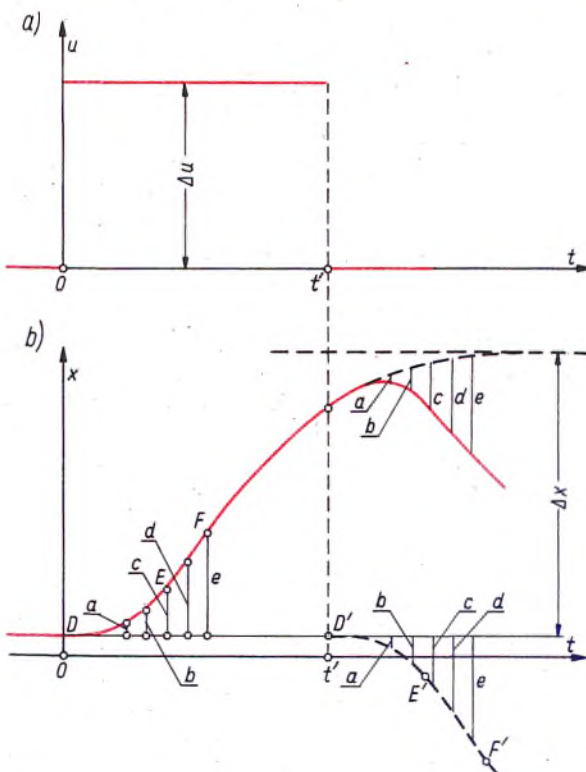
•4.46A• Charakterystyki skokowe typowych obiektów regulacji: a) wymuszenie; b) odpowiedź obiektu statycznego; c) odpowiedź obiektu astatycznego

rystyki dynamicznej regulatora od wartości częstotliwości krytycznej układu bez korekcji.

Stosując metody oparte na pomiarze charakterystyk skokowych zakłada się, że obiekt można uważać za obiekt inercyjny z opóźnieniem. Dla takiego obiektu określa się, na podstawie charakterystyki skokowej, podstawowe parametry, a następnie z tabel dobiera się wartości nastaw regulatora.

•4.46• Jakie parametry obiektów trzeba znać przy nastawianiu regulatorów według cech przebiegu przejściowego?

Charakterystyki skokowe typowych obiektów regulacji przedstawiono na rys. 4.46A. Charakterystyki takie otrzymuje się drogą rejestracji zmian wielkości wyjściowej x obiektu, po wprowadzeniu na jego wejście skokowej zmiany wielkości sterującej u . Jeżeli ze względów technologicznych nie można dopuścić do występowania w obiekcie długotrwałych odchyśleń od punktu pracy, to wtedy stosuje się wymuszenie w postaci impulsu prostokątnego (rys. 4.46B). Wymuszenie takie stanowi w rzeczywistości sumę dwóch jednakowych lecz przeciwnie skierowanych wymuszeń skokowych, występujących w chwilach $t = 0$ i $t = t'$. Na rysunku 4.46B przedstawiono graficzną metodę uzupełniania



•4.46B• Odpowiedź obiektu statycznego na wymuszenie impulsowe: a) wymuszenie; b) odpowiedź obiektu

„brakującej” części odpowiedzi skokowej obiektu, na przykładzie obiektu inercyjnego z opóźnieniem.

Na rysunku 4.46A przedstawiono charakterystyki dwóch podstawowych rodzajów obiektów: inercyjnego z opóźnieniem i całkującego (astatycznego) z opóźnieniem. Obiekty inercyjne posiadające charakterystykę skokową taką jak na rys. 4.46A (b) są opisywane, z przybliżeniem wystarczającym dla doboru nastaw regulatorów, transmitancją

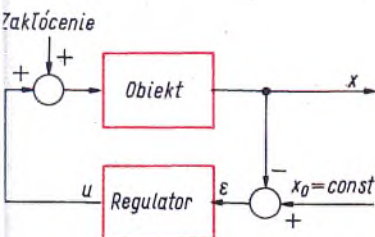
$$G(s) = k_{ob} \frac{e^{-sT_o}}{sT + 1}$$

Opóźnienie zastępcze T_o , zastępczą stałą czasową T oraz wzmacnienie obiektu k_{ob} określa się na podstawie charakterystyki skokowej w sposób przedstawiony na rysunku.

Obiekty astatyczne, posiadające charakterystyki skokowe takie jak na rys. 4.46A (c), opisuje z wystarczającą dokładnością transmitancja

$$G(s) = \frac{e^{-sT_o}}{sT_c}$$

Znając wartości opóźnień zastępczych, stałych czasowych i wzmacnień obiektów można, posługując się odpowiednimi tabelami, ustalać przybliżone wartości nastaw regulatorów.

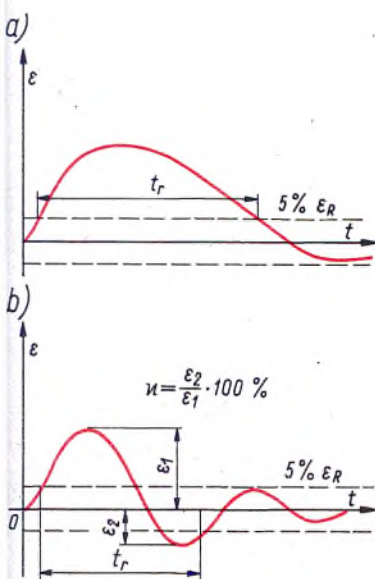


•4.47A• Schemat blokowy układu regulacji

•4.47• Jakie kryteria jakości stosuje się przy nastawianiu regulatorów według cech przebiegu przejściowego?

Zgodnie z nazwą metody doboru nastaw według cech przebiegu przejściowego podstawą do oceny jakości regulacji jest tu przebieg przejściowy otrzymywany w układzie z regulatorem. Najczęściej rozpatruje się przebieg uchybu przy zakłóceniu skokowym działającym tak jak pokazuje rys. 4.47A. Na podstawie otrzymanego w takich warunkach przebiegu uchybu (rys. 4.47B), określa się wartość współczynnika przeregulowania κ oraz czasu regulacji t_r ; ϵ_R oznacza uchyb, który powstałby, gdyby zakłócenie działało na obiekt bez regulatora.

Wartości współczynnika przeregulowania i czasu regulacji często traktuje się jako **wskaźniki jakości regulacji**. Na ogół dla układu regulacji podaje się tylko dopuszczalną wartość przeregulowania, żądając jednocześnie, aby przy tym przeregulowaniu czas regulacji był możliwie najkrótszy. Najczęściej spotyka się wartość dopuszczalnego przeregulowania w granicach od 0 do 5% lub 20%. W pierwszym przypadku otrzymany przebieg można uważać w praktyce za aperiodyczny. Przeregulowanie mieści się bowiem w pięcioprocentowym marginesie 5% ϵ_R (patrz rys. 4.47B) i czas regulacji liczy się do pierwszego „wejścia” wykresu uchybu do wewnątrz „paska” określonego tym marginesem. W drugim przypadku przeregulowania są wyraźne i sygnał uchybu parokrotnie zmienia znak zanim jego wartość trwale stanie



•4.47B• Zmiany uchybu regulacji po wprowadzeniu zakłócenia skokowego:
a) przeregulowanie mniejsze od 5%;
b) przeregulowanie powyżej 5%

4.48A. Wzory do obliczania nastaw dla regulatorów współpracujących z obiektami astatycznymi z opóźnieniem

Wskaźnik jakości regulacji Typ regulatora	$\kappa = 2...5\% \min t_r$	$\kappa = 20\% \min t_r$	$\min \int_0^{t_r} \varepsilon^2(\tau) d\tau$
P	$k_p \frac{T_0}{T_c} = 0,5$	$k_p \frac{T_0}{T_c} = 0,7$	$k_p \frac{T_0}{T_c} = 1$
PI	$k_p \frac{T_0}{T_c} = 0,5$ $T_i = 5T_0$	$k_p \frac{T_0}{T_c} = 0,7$ $T_i = 3T_0$	$k_p \frac{T_0}{T_c} = 1$ $T_i = 4,3T_0$
PID	$k_p \frac{T_0}{T_c} = 0,65$ $T_i = 5T_0$ $T_d = 0,23T_0$	$k_p \frac{T_0}{T_c} = 1,1$ $T_i = 2T_0$ $T_d = 0,37T_0$	$k_p \frac{T_0}{T_c} = 1,3$ $T_i = 1,6T_0$ $T_d = 0,5T_0$

T_c — stała czasowa całkowania obiektu, T_0 — czas opóźnienia obiektu, k_p — wzmacnienie regulatora, T_i — czas zdwojenia regulatora, T_d — stała czasowa różniczkowania regulatora (czas wyprzedzenia regulatora), t_r — czas regulacji.

4.48B. Wzory do obliczania nastaw dla regulatorów współpracujących z obiektami statycznymi inercyjnymi z opóźnieniem

Wskaźnik jakości regulacji Typ regulatora	$\kappa = 2...5\% \min t_r$	$\kappa = 20\% \min t_r$	$\min \int_0^{t_r} \varepsilon^2(\tau) d\tau$
P	$k_p k_0 \frac{T_0}{T} = 0,9$	$k_p k_0 \frac{T_0}{T} = 0,7$	
PI	$k_p k_0 \frac{T_0}{T} = 0,6$ $T_i = 0,8T_0 + 0,5T$	$k_p k_0 \frac{T_0}{T} = 0,7$ $T_i = T_0 + 0,3T$	$k_p k_0 \frac{T_0}{T} = 1$ $T_i = T_0 + 0,3T$
PID	$k_p k_0 \frac{T_0}{T} = 0,95$ $T_i = 2,4T_0$ $T_d = 0,4T_0$	$k_p k_0 \frac{T_0}{T} = 1,2$ $T_i = 2T_0$ $T_d = 0,4T_0$	$k_p k_0 \frac{T_0}{T} = 1,4$ $T_i = 1,3T_0$ $T_d = 0,5T_0$

T — stała czasowa obiektu, k_0 — wzmacnienie obiektu, T_0 — czas opóźnienia obiektu, k_p — wzmacnienie regulatora, T_i — czas zdwojenia regulatora, T_d — czas wyprzedzenia regulatora, t_r — czas regulacji.

się mniejsza od 5% ε_R . Do nastawiania regulatorów według cech przebiegu przejściowego stosuje się także kryteria tzw. całkowite, żądające minimalnej wartości całki modułu lub kwadratu uchybu. Kryteria całkowite są bardziej ogólne niż kryteria oparte na bezpośrednim pomiarze wartości prze-regulowania i czasu regulacji. Są one jednak stosowane dość rzadko, ponieważ w praktyce trudno jest wyznaczać wartości odpowiednich całek.

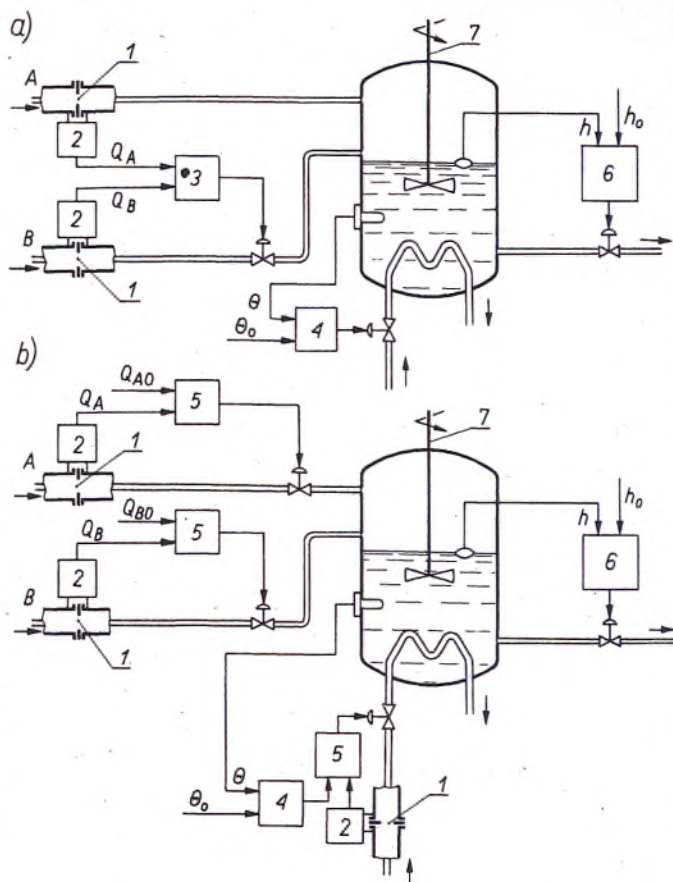
·4.48· Jakie są zależności pomiędzy parametrami obiektów a nastawami regulatorów?

Zależności pomiędzy wartościami parametrów obiektów a wartościami nastaw zapewniającymi uzyskanie założonych wskaźników regulacji podaje się w postaci tablic lub wykresów. Zależności takie dla trzech różnych wskaźników regulacji są zestawione w tabl. 4.48A i 4.48B.

·4.49· Czy przez właściwy dobór konfiguracji układu można poprawić jakość regulacji?

Oczywiście tak. W przemyśle istnieje wiele obiektów charakteryzujących się dużymi opóźnieniami i dużymi stałymi czasowymi. Jeżeli obiekty takie pracują przy dużych zakłóceniach, to mimo istnienia układów regulacji wahania wielkości regulowanej mogą być znaczne. Tam, gdzie jest to możliwe, staramy się więc eliminować zakłócenia przez regulację wielkości wejściowych obiektu. Jako przykład takiego działania na rys. 4.49 przedstawiono układy regulacji reaktora chemicznego. Do reaktora dostarczane są dwa składniki A i B o temperaturze zbliżonej do temperatury otoczenia. Reakcja powinna przebiegać w temperaturze 120°C. Do ogrzewania użyto nagrzewnicy zasilanej parą wodną. Reakcja jest endotermiczna czyli przebiega z poborem ciepła z otoczenia. Warunki technologiczne wymagają zachowania stałego stosunku wartości strumieni Q_A i Q_B obu składników, stałości temperatury oraz stałości poziomu w reaktorze.

Rysunek 4.49A przedstawia rozwiązanie, w którym do utrzymania stałej wartości trzech wymienionych wielkości służą dwa regulatory. Rozwiązanie to ma jednak wadę polegającą na tym, że wartości strumieni Q_A i Q_B mogą podlegać wahaniom, przy zachowaniu ich stałego stosunku. Warunki technologiczne wymagają wprowadzić tylko zachowania stałego stosunku strumieni Q_A i Q_B i regulator stosunku 3 zapewnia spełnienie tego wymagania. Jednak każda zmiana wartości strumieni Q_A i Q_B jest związana ze zmianą zapotrzebowania na ciepło niezbędne do ogrzania dostarczanych składników od temperatury ok. 20°C do temperatury 120°C oraz do przebiegu reakcji. Każda taka zmiana stanowi zakłócenie dla układu regulacji temperatury. Dlatego lepsze jest rozwiązanie przedstawione na rys. 4.49b, w którym zamiast regulacji stosunku obu strumieni Q_A i Q_B wprowadzono stabilizację ich wartości.



•4.49• Układy regulacji reaktora chemicznego

1 — zwężki pomiarowe; 2 — przetworniki pomiarowe; 3 — regulator stosunku; 4 — regulator temperatury; 5 — regulatory przepływu; 6 — regulatory poziomu; 7 — mieszadło; θ — temperatura obiektu; θ_0 — wartość zadana temperatury; Q_A , Q_B — natężenie przepływu; Q_{A0} , Q_{B0} — wartości zadane natężeń przepływów

Zakłóceniem dla układu regulacji temperatury w reaktorze chemicznym jest także zmiana ciśnienia, powodująca zmianę natężenia przepływu pary ogrzewającej reaktor. Nie można jednak zastosować układu stabilizacji natężenia przepływu pary, ponieważ w przypadku pojawienia się uchybu temperatury wywołanego działaniem innych zakłóceń (np. zmiany temperatury otoczenia) trzeba to natężenie zmieniać. W tej sytuacji najlepsze wyniki uzyskamy stosując rozwiązanie pokazane na rys. 4.49b, czyli tzw. układ regulacji kaskadowej.

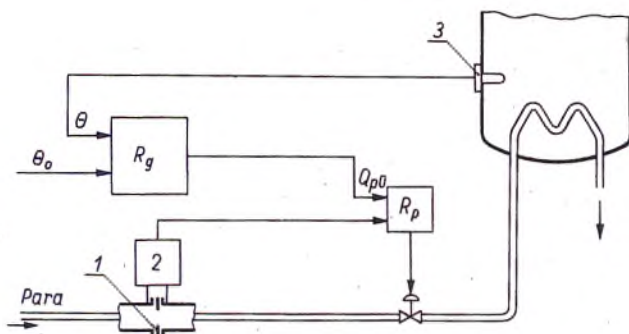
W opisie do rys. 4.49 podano jedynie, jaką funkcję spełniają poszczególne regulatory, nie określając ich rodzaju. Nie było to potrzebne, ponieważ regulatory są urządzeniami uniwersalnymi i w omawianym układzie można zastosować regulatory dowolnego typu dobierając ich parametry do właściwości dynamicznych obiektu.

•4.50• Co to jest regulacja kaskadowa?

Typowy przykład regulacji kaskadowej przedstawiono na rys. 4.50. Jest to układ regulacji temperatury w obiekcie

4.50 Kaskadowy układ regulacji temperatury

R_g — regulator główny; R_p — regulator pomocniczy; 1 — zwężka pomiarowa; 2 — przetwornik pomiarowy; 3 — termometr; θ — temperatura w obiekcie; θ_0 — temperatura zadana; Q_{p0} — żądana wartość natężenia przepływu pary; Q_p — natężenie przepływu pary



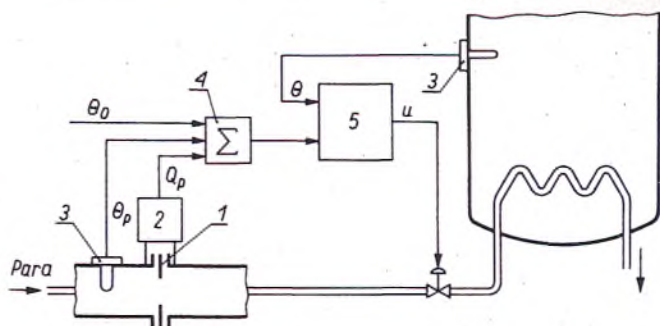
ogrzewanym parą. W układzie pracują dwa regulatory. Pierwszy z nich, zwany regulatorem głównym R_g , porównuje wartość temperatury w obiekcie z wartością zadaną. Regulator główny wytwarza wartość zadaną dla drugiego regulatora, zwanego regulatorem pomocniczym R_p . Regulator pomocniczy porównuje mierzoną wartość natężenia przepływu pary z wartością zadaną określoną sygnałem wyjściowym regulatora głównego i formuje sygnał oddziałujący na zawór regulujący przepływ pary. Cechą charakterystyczną układu regulacji kaskadowej jest to, że regulator główny nie oddziałuje bezpośrednio na zawór, lecz określa wartość wielkości pomocniczej (w omawianym przykładzie natężenia przepływu pary) potrzebną do utrzymania wartości zadanej wielkości regulowanej (w naszym przypadku temperatury). Dzięki takiemu rozwiązaniu zakłócenia wpływające na wartość wielkości pomocniczej (natężenia przepływu) są wyeliminowane, zanim spowodują powstanie znacznych odchyłek wielkości regulowanej. W przypadku takim jak opisany regulacja kaskadowa daje szczególnie dobre wyniki, ponieważ czas potrzebny do regulacji natężenia przepływu jest z reguły wielokrotnie mniejszy niż stała czasowa reaktora ogrzewanego parą. W układach regulacji kaskadowej najczęściej jako regulatory pomocnicze stosuje się regulatory typu P, a jako regulatory główne regulatory typu PI lub PID.

4.51 Czy mierząc zakłócenia działające na obiekt można poprawić rezultaty działania układu regulacji?

Tak. Jeżeli na obiekt działają zakłócenia, które można zmierzyć, ale których nie można wyeliminować, to wynik pomiaru zakłóceń można zastosować do „uprzedzenia” układu o pojawieniu się zakłóceń. Realizuje się to przez wprowadzenie do regulatora dodatkowego sygnału z układu pomiaru zakłóceń. Przykład układu regulacji z pomiarem zakłóceń przedstawia rys. 4.51. Jest to układ regulacji temperatury w obiekcie ogrzewanym parą wodną. Tym obiektem może być reaktor chemiczny, wymiennik ciepła itp. Przyjmijmy, że głównymi zakłóceniami działającymi na obiekt są wahania ciśnienia i temperatury pary. Układ regulacji

•4.51• Układ regulacji z pomiarem wartości zakłóceń

1 — zwężka pomiarowa; 2 — przetwornik pomiarowy; 3 — termometr; 4 — układ sumujący; 5 — regulator; θ_0 — temperatura zadana; θ — temperatura obiektu; θ_p — temperatura pary



kaskadowej eliminuje wahania ciśnienia, ale pozostają wahania temperatury. Jeżeli wahania te są duże, a zależy nam na dokładności regulacji temperatury w obiekcie, to można temperaturę pary zmierzyć i zmieniać jej dopływ tak, aby ilość ciepła dostarczana do obiektu pozostawała niezmieniona.

Możliwe są różne rozwiązania występującego tu zadania regulacji — jedno z nich przedstawiono na rys. 4.51. Mierzone jest natężenie przepływu i temperatura pary dostarczanej do obiektu, a wyniki tych pomiarów oddziałują na regulator sterujący dopływ pary. Dla skompensowania wpływu wahań temperatury i natężenia przepływu pary na temperaturę w obiekcie, regulator powinien realizować zależność

$$u = f[(\theta_0 + k_1 \theta_p + k_2 Q_p) - \theta]$$

W wyrażeniu tym θ_0 jest sygnałem wartości zadanej, θ — sygnałem temperatury w obiekcie, k_1 i k_2 są współczynnikami proporcjonalności, a funkcja f symbolizuje przekształcenia wykonywane w regulatorze — jej postać zależy od tego, czy zastosowano regulator P, PI czy PID. Współczynniki k_1 i k_2 powinny być dobrane odpowiednio do wrażliwości temperatury w obiekcie na zmiany temperatury i natężenia przepływu pary.

•5.1• Jakiego rodzaju powody stosowania urządzeń cyfrowych w układach automatyki?

O szerokim zastosowaniu urządzeń cyfrowych w automatyce zadecydował, z jednej strony, rozwój techniki cyfrowej i związane z nim obniżenie kosztów i poprawa parametrów urządzeń cyfrowych. Z drugiej strony zastosowanie techniki cyfrowej w układach automatyki stwarza szereg nowych możliwości. Jako najważniejsze z nich można wymienić: — Operatywna kontrola wartości wielu parametrów, szybkiego wykrywania stanów awaryjnych, rejestracji warunków w jakich przebiega proces technologiczny. Spotykane, przy sterowaniu przebiegu procesów w dużych obiektach, przypadki w których operator kierujący procesem powinien śledzić wartości kilkuset parametrów wymagają stosowania nowych rozwiązań technicznych.

— Realizacja skomplikowanych reguł sterowania. Przy stosowaniu standardowej analogowej aparatury regulacyjnej budowa układów bardziej skomplikowanych niż układy regulacji kaskadowej lub układy z pomiarem zakłóceń, stwarza bardzo duże trudności. Zastosowanie urządzeń cyfrowych, a w szczególności komputerów, umożliwia realizację praktycznie dowolnie złożonych algorytmów sterowania, ze sterowaniem optymalizacyjnym i adaptacyjnym łącznie.

— Przyspieszanie procesów przebiegających etapowo, w których przejście do następnego etapu może nastąpić dopiero po osiągnięciu pewnego stanu, wykrywanego przez kontrolę wartości wielu parametrów. Tak np. zastosowanie komputerów do rozruchu i odstawiania jednostek energetycznych dużej mocy znacznie skraca czasy tych operacji i zmniejsza niebezpieczeństwo awarii.

— Zwiększenie dokładności sterowania.

— Automatyzacja obiegu materiałów w zakładzie, dzięki czemu można realizować indywidualne zamówienia nie tracąc korzyści, jakie daje produkcja seryjna.

•5.2• Jakiego rodzaju urządzenia cyfrowe stosuje się w układach automatyki?

Urządzenia cyfrowe można podzielić na dwie grupy: **uniwersalne i specjalizowane**. Urządzenia uniwersalne, czyli komputery, charakteryzują się tym, że program ich działania jest umieszczany w ich pamięci i może być łatwo zmieniany przez zmianę zawartości pamięci. Program działania urządzeń specjalizowanych jest określany w okresie ich projektowania i budowy, jeżeli więc przy projektowaniu nie przewidziano zmian programu to potem już tego w zasadzie zrobić nie można.

W trakcie rozwoju techniki cyfrowej proporcje między urządzeniami uniwersalnymi i specjalizowanymi zmieniały się na korzyść tych pierwszych: wszystkie urządzenia cyfrowe taniały, ale komputery taniały szybciej niż urządzenia specjalizowane. Wraz ze zmianami cen zmieniał się zakres zastosowań: im tańsze są komputery tym mniejszy jest zakres zastosowań urządzeń specjalizowanych. Obecnie układ cen jest taki, że jednostka centralna małego prostego komputera (tzw. mikroprocesor) kosztuje kilkanaście do kilkudziesięciu dolarów, a płytką z zakresu wyposażenia specjalnego, zawierającą kilkanaście elementów, kosztuje kilkaset dolarów.

Układ cen oraz takie zalety układów z komputerami jak elastyczność, łatwość rozbudowy i łatwość wprowadzenia zmian w trakcie rozruchu i eksploatacji układu sterowania, spowodowały ograniczenie zastosowań urządzeń specjalizowanych do przypadków bardzo prostych i powtarzalnych. Urządzenie specjalizowane stosuje się także tam, gdzie jest wymagana bardzo duża szybkość (np. w niektórych interpolatorach w układach sterowania programowego).

W układach automatyki stosuje się komputery uniwersalne, tzn. takie które mogą wykonywać również inne zadania, jak np. obliczenia naukowe. Nie ma obecnie na rynku komputerów produkowanych z przeznaczeniem dla celów sterowania. Najczęściej stosowane są tu minikomputery i mikrokomputery.

·5.3· Jakiego rodzaju zadania są realizowane przez urządzenia cyfrowe w układach automatyki?

Zadania spełniane przez urządzenia cyfrowe w układach automatyki można ogólnie sklasyfikować następująco:

— Gromadzenie i obróbka informacji o procesie sterowanym, czyli tzw. centralna rejestracja i przetwarzanie danych (w skrócie CRPD).

— Bezpośrednie sterowanie cyfrowe przebiegu procesu (w skrócie BSC lub DDC — od angielskiego terminu Direct Digital Control).

— Pośrednie sterowanie przebiegu procesu realizowane za pomocą regulatorów współpracujących z komputerami. Takie sterowanie jest nazywane **sterowaniem nadrzędnym**.

— Sterowanie włączenia i wyłączenia silników, zaworów i innych mechanizmów, sterowanie transportu, magazynów itp.

Tego typu sterowanie nazywa się **sterowaniem sekwencyjnym**.

— Sterowanie przesunięć liniowych i kątowych. Typowymi przykładami są tu układy sterowania programowego obrabiarek, układy sterowania walców zgniataczy i in.

•5.4• Jakie operacje wykonuje się w układach centralnej rejestracji i przetwarzania danych (CRPD)?

Układy pomiarowe, nazywane układami centralnej rejestracji i przetwarzania danych (CRPD), powstały dzięki zastosowaniu techniki cyfrowej do sterowania automatycznego. Funkcje spełniane przez te układy można podzielić następująco:

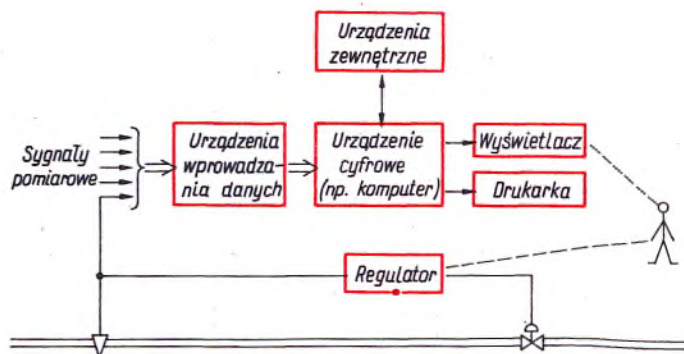
— Obieganie w określonej kolejności wielu (kilkudziesięciu do kilkuset) kanałów pomiarowych, dokonywanie pomiarów cyfrowych z przełączaniem zakresów pomiaru i ewentualnie linearyzacja nieliniowości charakterystyk źródeł sygnałów, cyfrowa rejestracja wyników pomiarów na papierze (druk) oraz na taśmie papierowej lub magnetycznej.

— Wykrywanie przekroczeń założonych ograniczeń zmian wielkości mierzonych, sygnalizacja zaistnienia przekroczenia, drukowanie wartości przekraczających wartości graniczne w odmienny sposób, tzn. kolorem czerwonym, ze specjalnym znakiem, na oddzielnej drukarce lub tp.

— Wykonywanie prostych obliczeń koniecznych do przedstawiania wyników pomiarów w postaci dogodnej dla operatora, jak np. obliczanie wartości wielkości interesującej operatora, trudnej lub niemożliwej do zmierzenia bezpośrednio, związanej znanymi zależnościami matematycznymi z innymi łatwymi do zmierzenia wielkościami fizycznymi, wyznaczanie wartości maksymalnej lub minimalnej parametru, tendencji zmian wartości parametrów i in.

— Przetwarzanie zbieranych danych wg programów umożliwiających uzyskanie uogólnionych, syntetycznych wskaźników analizowanego procesu.

Dwie pierwsze spośród wymienionych funkcji są nieodłącznie związane z definicją centralnej rejestracji i są realizowane w zasadzie przez wszystkie układy obejmowane tą nazwą. Pozostałe operacje mogą być wykonywane lub nie wykonywane, zależnie od wyposażenia układu w sprzęt i oprogramowanie.



•5.4• Schemat strukturalny centralnej rejestracji i przetwarzania danych

W układach CRPD informacje wejściowe są dostarczane do obiektu (rys. 5.4) a informacje wyjściowe, przeznaczone wyłącznie dla obsługi są wyprowadzane do urządzeń wskazujących i rejestrujących.

·5.5· Gdzie są stosowane układy centralnej rejestracji i przetwarzania danych (CRPD)?

Układy CRPD umożliwiają znaczne odciążenie personelu kontrolującego przebieg procesu technologicznego lub eksperymentu naukowego. Potrzeba automatyzacji zbierania i porządkowania wyników pomiarów oraz rozwój techniki cyfrowej doprowadziły do wielkiego rozpowszechnienia tych układów. Jako najbardziej typowe zastosowania można wymienić:

— Kontrola przebiegu procesu technologicznego; instalowane w tym celu urządzenia zastępują wiele rejestratorów i wskaźników analogowych, sygnalizują stany alarmowe, mogą wykonywać obliczenia syntetycznych wskaźników charakteryzujących kontrolowany proces.

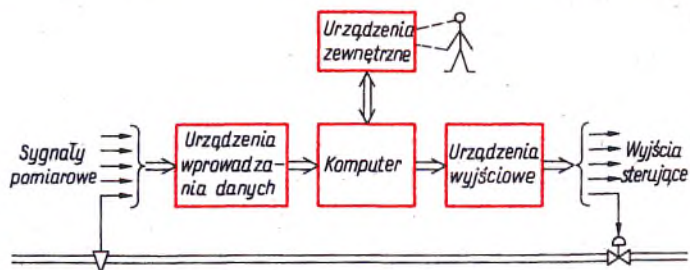
— Kontrola międzyoperacyjna oraz kontrola końcowa i wykonywanie dokumentów kontroli (atestów); np. w przemyśle teletechnicznym przy produkcji podzespołów (rezystorów, kondensatorów itp.), w przemyśle hutniczym do ciągłej kontroli jakości blach i drutów, w przemyśle szklarskim, farmaceutycznym itd.

— Zbieranie danych dla identyfikacji procesu; do tego celu są instalowane proste urządzenia dokonujące cyfrowych pomiarów wartości wielu parametrów i rejestrujące wyniki tych pomiarów na taśmie papierowej lub magnetycznej; wyniki pomiarów są później przetwarzane przez komputer w ośrodku obliczeniowym.

— Zbieranie, rejestracja i przetwarzanie danych — wyników eksperymentów naukowych; zastosowanie techniki cyfrowej znacznie przyspieszyło i ułatwiło prace pomiarowe; są tu stosowane zarówno proste układy dokonujące pomiarów i tabelaryzacji wyników jak również rozbudowane układy zawierające komputer, mogące dostarczać wyniki badań w postaci syntetycznej, np. w postaci wykresów; liczba kanałów pomiarowych waha się od kilkudziesięciu do kilkuset; nie występuje tu tak ostro jak w układach przemysłowych problem zakłóceń i odporności na zmienne warunki zewnętrzne; zakres zastosowań układów CRPD w badaniach naukowych i technicznych jest bardzo szeroki; są one stosowane w badaniach niszczących i nieniszczących konstrukcji budowlanych, statków, samolotów, w badaniach tunełowych, w analizie chemicznej, w ruchomych laboratoriach i w wielu innych dziedzinach; można powiedzieć, że nowoczesnych badań naukowych i technicznych nie można prowadzić bez układów CRD i CRPD.

— Na statkach — pomiar i rejestracja podstawowych parametrów silnika głównego i innych ważniejszych urządzeń (np. chłodni).

•5.6• Schemat strukturalny bezpośredniego sterowania cyfrowego



— W medycynie — kontrola stanu chorego (lub chorych) przez pomiar, analizę i sygnalizację przekroczeń zadanych wartości parametrów charakteryzujących ten stan.

•5.6• Jakiego rodzaju sterowanie nazywamy „bezpośrednim sterowaniem cyfrowym”?

Termin „bezpośrednie sterowanie cyfrowe” oznacza takie sterowanie procesu przez komputer, przy którym komputer oddziałuje bezpośrednio na proces objęty sterowaniem. Oznacza to, że komputer bezpośrednio (rys. 5.6) oddziałuje na elementy wykonawcze: zawory, kłapy itp.

W układach bezpośredniego sterowania cyfrowego występuje z reguły:

— Wiele wejść analogowych przyłączonych przez przetworniki a/c.

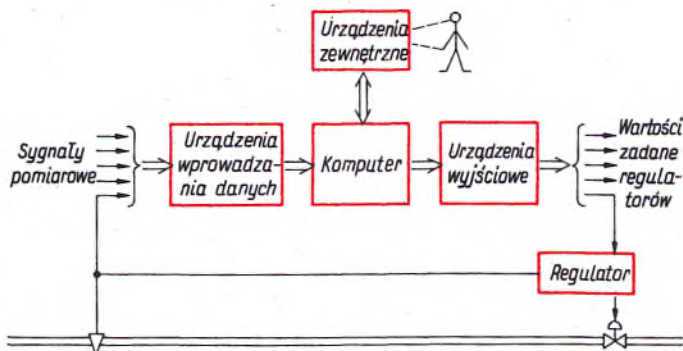
— Jeden centralny układ cyfrowy (obecnie jest to zawsze komputer) obsługujący wiele kanałów regulacyjnych, pracujący z podziałem czasu pomiędzy poszczególne kanały.

— Wiele kanałów wyjściowych oddziałujących bezpośrednio na proces. Istnienie tych kanałów wyjściowych stanowi najistotniejszą cechę BSC. Liczba kanałów wyjściowych jest jednak zawsze mniejsza niż liczba kanałów wejściowych.

— Stosunkowo proste algorytmy obliczania wartości sygnałów wyjściowych. Najczęściej są to algorytmy typu PID.

•5.7• Co to jest pośrednie sterowanie cyfrowe?

Pośrednim sterowaniem cyfrowym jest nazywany taki system sterowania procesu przez komputer, w którym komputer



•5.7• Schemat pośredniego sterowania cyfrowego

steruje zmianami nastaw regulatorów analogowych PID (rys. 5.7). Dla takiego sposobu sterowania jest używany również termin **sterowanie nadrzędne**. Ten drugi termin podkreśla fakt, że komputer spełnia tu funkcje nadrzędne realizując program wyższego rzędu, np. optymalizacji według określonego kryterium.

5.8. Jakie różnice występują przy realizacji układów bezpośredniego i pośredniego sterowania cyfrowego?

Problemy techniczne występujące przy realizacji systemów bezpośredniego i pośredniego sterowania cyfrowego są podobne. Urządzenia wejściowe wprowadzające do komputera sygnały od obiektu są takie same, niezależnie od realizowanego programu sterowania. Po stronie wyjściowej w układach BSC sygnały z komputera są używane do sterowania urządzeń wykonawczych, a w układach sterowania pośredniego jako wartości zadane regulatorów analogowych. Jednak postać, do której należy przetwarzać otrzymywane z komputera sygnały cyfrowe, jest w obu przypadkach taka sama. Również żądana dokładność formowania sygnałów jest w obu przypadkach jednakowa. Szczególnie wyraźnie widać to w przypadku stosowania aparatury pneumatycznej; standardowy sygnał 0,2 ... 1 atn jest wymagany zarówno przez urządzenia wykonawcze jak i przez regulatory analogowe. Do sterowania urządzeń wykonawczych jest potrzebny tylko sygnał o większej mocy. W systemach pośredniego sterowania cyfrowego używa się regulatorów przystosowanych do współpracy z komputerem, umożliwiających ustawianie wartości zadanej sygnałami z komputera. Regulatory takie, stosuje się również w układach BSC jako regulatory awaryjne.

Istotne różnice występują w zabezpieczeniach awaryjnych stosowanych w układach bezpośredniego i pośredniego sterowania cyfrowego w przypadku awarii komputera. Przy sterowaniu pośrednim problem sprowadza się do natychmiastowego wykrycia uszkodzenia i niedopuszczenia do wykonania błędnego polecenia. W układach, w których proces jest prowadzony blisko ograniczeń awaryjnych, w razie awarii komputera ustawia się ustalone wcześniej „bezpieczne” wartości zadane dla wszystkich regulatorów.

W układach BSC, mimo stałego wzrostu niezawodności komputerów, ważniejsze pętle regulacji są zabezpieczone przez instalowanie rezerwowej aparatury analogowej: stacyjek sterowania ręcznego oraz regulatorów PI lub PID. Urządzenia te mogą być włączone w pętle regulacyjne przez operatora lub automatycznie. Na ogół zabezpieczenia wymaga 10 ... 40% pętli regulacyjnych. Konieczność instalowania urządzeń rezerwowych zwiększa koszty układów BSC i obniża ich konkurencyjność w stosunku do układów z regulatorami konwencjonalnymi, nie wymagających stosowania tego typu urządzeń.

Instalowanie rezerwowej aparatury analogowej nie zawsze

daje wystarczająco dobre rezultaty. Dotyczy to przede wszystkim systemów, w których jest duża liczba wejść i wyjść cyfrowych i gdzie musi być ściśle zachowana kolejność wykonywania wielu czynności. W tych przypadkach oraz wówczas, gdy nie pociąga to za sobą wyraźnego zwiększenia kosztów, sterowanie awaryjne powierza się drugiemu komputerowi. Według materiałów firmy Faxboro niezawodność systemu zwiększa się dzięki temu ok. 100 razy.

·5.9· Z jakich elementów buduje się urządzenie cyfrowe?

Większe urządzenia cyfrowe buduje się z podstawowych, elementarnych układów cyfrowych. Układy te nazywa się **elementami** lub **modułami cyfrowymi**. Zasadniczą ich cechą jest praca przy sygnałach przyjmujących wartość „zero” lub „jeden” oraz standaryzacja poziomów i mocy sygnałów, umożliwiającą współpracę wielu elementów w różnych wzajemnych połączeniach. Pełny zestaw elementów logicznych (cyfrowych), z których można zbudować dowolne urządzenie cyfrowe powinien zawierać:

- Elementy logiczne umożliwiające realizację dowolnej funkcji przełączającej. Warunek ten jest spełniony jeżeli w zestawie znajduje się element NOR lub NAND.

- Elementy pamięciowe umożliwiające realizację układów sekwencyjnych.

- Elementy czasowe.

- Elementy wejściowe i wyjściowe umożliwiające łączenie układu cyfrowego z urządzeniami zewnętrznymi.

Elementy cyfrowe można budować korzystając z rozmaitych procesów fizycznych. Obecnie stosowane są elementy półprzewodnikowe i pneumatyczne. Wśród elementów półprzewodnikowych najbardziej rozpowszechnione są układy scalone wykonywane techniką TTL. Układy scalone TTL wyparły wszystkie inne rodzaje elementów z komputerów, kariera ich jest spowodowana głównie opracowaniem układów o dużym i średnim stopniu scalenia — zawierających w jednej obudowie całe bloki funkcjonalne. W układach automatyki znajdują jeszcze niekiedy zastosowanie elementy cyfrowe budowane z oddzielnych tranzystorów i diod. Elementy takie są znacznie droższe, ale są bardziej odporne na zakłócenia. W układach automatyki przemysłowej są stosowane także elementy wykonywane techniką DZTL, charakteryzującą się znacznie zwiększoną odpornością na zakłócenia.

Elementy pneumatyczne mogą być z częściami ruchomymi (np. membranowe) lub bez części ruchomych, z wykorzystaniem wzajemnego oddziaływania strumienia powietrza. W praktyce, jak dotąd, większe zastosowanie znalazły elementy z częściami ruchomymi, przede wszystkim membranowe. Z pneumatycznych elementów cyfrowych są budowane układy sterujące, zabezpieczające i blokady czyli typowe układy przekazywnikowe. Elementów tych nie stosuje się do budowy bardziej skomplikowanych układów cyfrowych.

·5.10· Skąd pochodzi nazwa „elementy logiczne” nadawana elementom cyfrowym?

Elementy, z których buduje się układy cyfrowe są nazywane elementami cyfrowymi albo elementami logicznymi. Nazwa elementy logiczne pochodzi stąd, że służą one do realizacji funkcji logicznych, czyli funkcji zapisywanych przy użyciu algebry Boole'a, operujących sygnałami dwuwartościowymi. Najważniejsze funkcje logiczne to:

— Suma logiczna, zapisywana w postaci

$$y = x_1 + x_2 + x_3 + \dots$$

przy czym sygnał wyjściowy y oraz sygnały wejściowe x_1, x_2, x_3 przyjmują wartości „zero” lub „jeden”. Funkcja sumy jest równa 1 jeżeli przynajmniej jeden z sygnałów wejściowych ma wartość 1. Inaczej mówiąc sygnał wyjściowy $x = 1$ jeżeli $x_1 = 1$ lub $x_2 = 1$ lub $x_3 = 1 \dots$ itd. Elementy realizujące funkcję sumy logicznej są nazywane często elementami LUB (OR).

— Iloczyn logiczny, zapisywany w postaci

$$y = x_1 x_2 x_3 \dots$$

Funkcja ta ma wartość 1 tylko wtedy, kiedy wszystkie sygnały wejściowe są równe 1. Elementy realizujące funkcję iloczynu logicznego nazywa się elementami I (AND)

— Negacja (zaprzeczenie)

$$y = \bar{x}$$

Negacja działa w ten sposób, że $y = 1$, gdy $x = 0$ oraz $y = 0$ gdy $x = 1$. Elementy realizujące funkcję negacji nazywa się elementami NIE (NOT).

·5.11· Jakie funkcje logiczne realizują elementy NAND i NOR?

Element NAND, którego nazwa pochodzi od NOT AND — realizuje zależność

$$y = \overline{x_1 x_2 x_3 \dots}$$

Jest to negacja iloczynu. Sygnał wyjściowy y ma wartość 0 tylko wtedy, kiedy iloczyn sygnałów wejściowych jest równy 1, tzn. gdy wszystkie sygnały wejściowe $x_1, x_2, x_3 \dots$ są równe 1.

Element NOR, którego nazwa pochodzi od NOT OR, realizuje zależność

$$y = \overline{x_1 + x_2 + x_3 + \dots}$$

Jest to negacja sumy. Sygnał wyjściowy y jest równy 1 tylko wtedy, kiedy suma sygnałów wejściowych jest równa 0, tzn. gdy wszystkie sygnały wejściowe mają wartość 0.

Zarówno elementy NAND jak elementy NOR pozwalają zrealizować dowolną funkcję przełączającą. Wynika to z praw de Morgana stwierdzających, że mając jeden z tych

elementów można zrealizować zarówno sumę, jak też iloczyn logiczny. Prawa de Morgana

$$\overline{x_1 x_2} = \overline{x_1} + \overline{x_2}$$

$$\overline{x_1 + x_2} = \overline{x_1} \overline{x_2}$$

można udowodnić podstawiając do podanych zależności wszystkie możliwe wartości argumentów.

·5.12· Jakich układów scalonych TTL używa się do budowy urządzeń cyfrowych?

Układy scalone TTL są obecnie najpowszechniej stosowanymi układami komputerów i cyfrowych urządzeń automatyki. Można wśród nich wyodrębnić trzy grupy różniące się stopniem scalenia. Pozornie jest to podział z punktu widzenia technologii wytwarzania, w rzeczywistości jest on ważny dla użytkownika. Grupy o których mowa są następujące:

— Elementy logiczne takie jak elementy kombinacyjne (NAND, AND, NOR) i przerzutniki. Są to układy w małym stopniu scalenia, zawierające nie więcej niż kilkanaście elementów takich jak diody i tranzystory.

— Bloki funkcjonalne takie jak licznik, rejestr, komutator, konwerter kodu itp. Są to układy o średnim stopniu scalenia, równoważne układom zawierającym kilkadziesiąt tradycyjnych elementów.

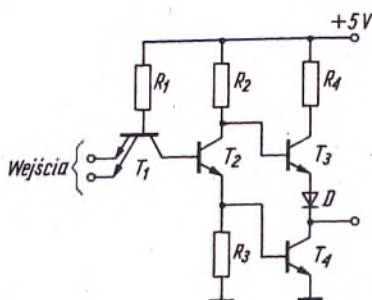
— Zespoły funkcjonalne, takie jak mikroprocesor, pamięć czy przetwornik analogowo-cyfrowy. Są to układy o dużym stopniu scalenia, równoważne układom zawierającym kilkadziesiąt elementów.

W urządzeniu cyfrowym zawierającym bloki funkcjonalne takie jak liczniki czy rejestry, można te bloki zmontować z elementów logicznych o małym stopniu scalenia lub też zastosować typowe bloki o średnim stopniu scalenia. Rozwiązanie drugie jest kilkakrotnie tańsze, a ponadto urządzenie będzie bardziej odporne na zakłócenia (dzięki mniejszej liczbie połączeń) i będzie pobierało mniejszy prąd zasilania. Stosowane powinny być więc układy o dużym stopniu scalenia. W Polsce produkuje się obecnie elementy TTL małej skali oraz niepełny jeszcze asortyment o średnim stopniu scalenia.

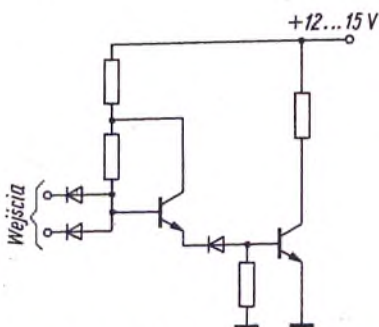
·5.13· Jaki jest podstawowy element techniki TTL?

Technika TTL (Transistor-Transistor Logic) jest obecnie najpopularniejszą techniką realizacyjną układów cyfrowych. Podstawowym elementem logicznym stosowanym w tej technice jest element NAND. Podstawowe parametry tego elementu są następujące:

- napięcie zasilania 5 V
- maksymalna liczba wejść 8
- obciążalność 10
- średni czas propagacji sygnału 10 ns
- straty mocy 10 mW



•5.13• Dwuwejściowy element NAND TTL



•5.14• Dwuwejściowy element NAND DZTL

- sygnał 0: typowy — 0,2 V, maksymalny — 0,4 V
- sygnał 1: typowy — 3,3 V, minimalny — 2,4 V
- margines zakłóceń: typowy — 1 V, minimalny — 0,4 V

Schemat elementu NAND jest przedstawiony na rys. 5.13. Jeśli przynajmniej na jedno wejście elementu jest podany sygnał 0, to tranzystor T_1 jest nasycony, a tranzystor T_2 zatkany. Wobec tego rezystor R_3 podając na bazę tranzystora T_4 potencjał ziemi powoduje jego zatkanie. Tranzystor T_3 pracuje wtedy jako wtórnik emiterowy, napięcie wyjściowe elementu jest równe ok. 3,3 V i mało zależy od obciążenia.

W układach o małym stopniu scalenia w jednej kostce znajduje się od jednego (ośmiowejściowego) do czterech (dwuwejściowych) elementów NAND.

•5.14• Jakie właściwości mają elementy wykonane techniką DZTL?

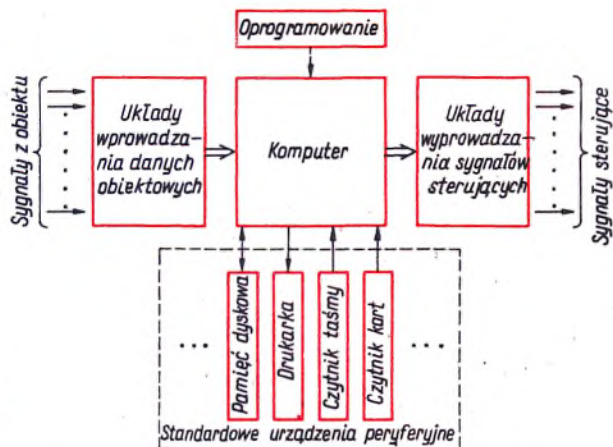
Nazwa techniki jest skrótem pełnej nazwy w języku angielskim: Diode-Zener Transistor Logic, w której wymieniono najważniejsze używane elementy półprzewodnikowe. Jedną z możliwych realizacji elementu DZTL jest przedstawiona na rys. 5.14. Jest to element NAND (negacja iloczynu). Elementy DZTL charakteryzują się zwiększonym (w stosunku do elementów TTL) napięciem zasilania (12 ... 15 V) oraz bardzo dużym marginesem zakłóceń. Uzyskano to dzięki zwiększonemu poborowi mocy i wydłużonemu czasowi propagacji sygnału.

Wydłużenie czasu propagacji, jest w wielu przypadkach czynnikiem pozytywnym, zwiększającym dynamiczną odporność na zakłócenia. Dalsze zwiększanie tego czasu można uzyskać przez dołączenie zewnętrznych kondensatorów spowalniających. Wiele firm produkuje elementy cyfrowe przeznaczone do układów automatyki z wyprowadzeniami umożliwiającymi tego rodzaju spowalnianie. Spowalnianie elementu jest bowiem równoznaczne z filtracją wielu zakłóceń, a w układach automatyki bardzo duża szybkość nie jest na ogół potrzebna.

•5.15• Co jest potrzebne do realizacji komputerowego systemu sterowania?

Jak wynika z odpowiedzi na pytanie 5.2 bardziej rozbudowane układy CRPD oraz układy bezpośredniego i pośredniego sterowania cyfrowego są realizowane obecnie przy użyciu komputerów. Jednak sam komputer nie wystarcza dla zrealizowania układu sterowania. Na rys. 5.15 przedstawiono niezbędny zestaw urządzeń potrzebnych do realizacji komputerowego systemu sterowania.

Z dziedziny sprzętu znajdują się tam, obok komputera i jego standardowego wyposażenia w postaci urządzeń wejścia-wyjścia oraz pamięci zewnętrznych, także niestandardowe urządzenia wejścia-wyjścia. Urządzenia te są potrzebne do



•5.15• Zestaw urządzeń potrzebnych do realizacji komputerowego systemu sterowania

wprowadzania do komputera danych z obiektu oraz do wyprowadzania sygnałów sterujących obiekt.

Obok sprzętu do realizacji sterowania niezbędne jest oprogramowanie specjalistyczne, które z uniwersalnego urządzenia jakim jest komputer uczyni urządzenie wyspecjalizowane, przygotowane do realizacji konkretnego zadania. Oprogramowanie jest nową, nie spotykaną przy stosowaniu innych urządzeń, częścią układu sterowania.

Interesujące są koszty realizacji komputerowego układu sterowania. Otóż, według danych z literatury, koszt projektu i oprogramowania pochłania ok. 50% całych kosztów. Pozostałe 50% dzieli się mniej więcej równo na koszty komputera z urządzeniami standardowymi oraz na koszty wyposażenia specjalnego (niestandardowe urządzenie wejścia-wyjścia, pulpit operatora itp.).

•5.16• W jaki sposób jedno centralne urządzenie cyfrowe współpracujące z wieloma kanałami regulacyjnymi i pomiarowymi?

Urządzenia cyfrowe obsługujące wiele kanałów pomiarowych i regulacyjnych pracują z podziałem czasu pomiędzy poszczególne kanały. Urządzenie wybiera spośród wielu kanałów ten, z którym ma w danej chwili współpracować i wysyła jego adres do układu wprowadzania danych. Komutator przyłącza wybrany kanał do urządzenia. Następuje szybkie pobranie informacji i można przyłączyć kanał następny.

Przyłączanie kanałów może się odbywać według określonej kolejności i w określonym rytmie czasowym albo w zależności od stanu obiektu. Na ogół część sygnałów jest wprowadzana w stałym rytmie czasowym, a część zależnie od istniejącej sytuacji. Tak np. sygnały analogowe z reguły wprowadzane są w stałym rytmie, a niektóre sygnały dwuwartościowe wtedy, kiedy ich wartość ulega zmianie. Częstość przyłączania poszczególnych kanałów jest dostosowywana

do szybkości zmian sygnałów, czyli do stałych czasowych obiektów. W praktyce przemysłowej najczęściej (ok. raz na sekundę) są przyłączane układy pomiaru przepływów, a najrzadziej (raz na kilkadziesiąt sekund) układy pomiaru temperatur w dużych obiektach. Wprowadzenie informacji do urządzenia cyfrowego trwa kilkanaście do kilkudziesięciu mikrosekund, jest więc dosyć czasu na kolejną obsługę wielu kanałów.

•5.17. Jakiego rodzaju sygnały są przesyłane od obiektu do urządzenia cyfrowego?

Sygnały wyjściowe czujników stosowanych w przemyśle mają najczęściej postać **analogową**. Są to napięcia, prądy lub ciśnienia, których wartości zmieniają się proporcjonalnie do zmian mierzonego parametru. Najwygodniejszy do dalszej obróbki jest **stałoprądowy sygnał napięciowy**; sygnały występujące w innej postaci są więc przetwarzane na napięcia stałe. Poziom otrzymanego sygnału napięciowego w sposób istotny wpływa na trudności napotykane przy dalszej jego obróbce. W praktyce dużą grupę stanowią sygnały na poziomie kilkudziesięciu miliwoltów, do tej grupy należą np. sygnały wyjściowe termoelementów. Sygnały o poziomie kilku do kilkunastu woltów pochodzą od przetworników ciśnienie-napięcie, od układów pomiaru wielkości elektrycznych (prądy, napięcia, moce) itp.

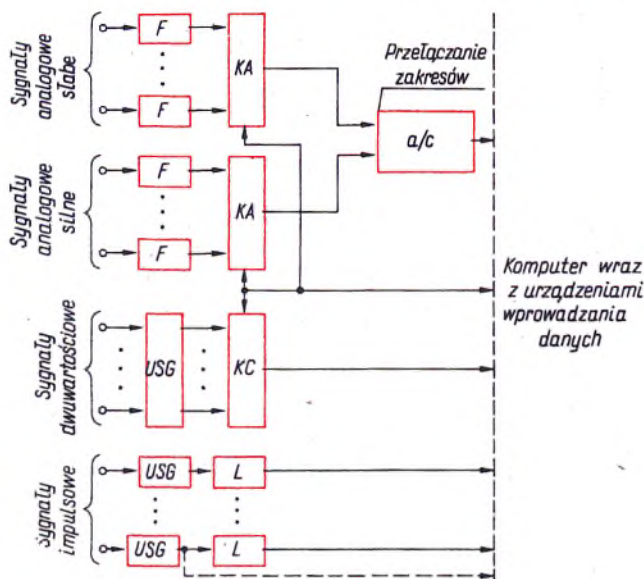
Osobną grupę stanowią sygnały o postaci **cyfrowej**. Są to przede wszystkim sygnały dwuwartościowe pochodzące od wyłączników sygnalizujących otwieranie i zamykanie zaworów, włączanie i wyłączanie silników itp. Sygnały cyfrowe mogą być też wytwarzane przez układy pomiarowe działające na zasadzie zliczania impulsów jak np. układy do pomiaru prędkości obrotowej, przepływu itp.

Wszystkie sygnały wejściowe wprowadzane do urządzeń sterowania cyfrowego muszą być uporządkowane i przekształcone do postaci odpowiedniej dla tego urządzenia tj. do postaci cyfrowej przedstawionej w odpowiednim kodzie; zadania te są realizowane przez układy **wprowadzania danych obiektowych**.

•5.18. Jakich urządzeń używa się przy wprowadzaniu sygnałów obiektów przemysłowych do urządzeń cyfrowych?

Schemat blokowy układu wprowadzania danych obiektowych do urządzenia cyfrowego przedstawiono na rys. 5.18. Układ ten umożliwia wprowadzanie wszystkich spotykanych w praktyce rodzajów sygnałów informujących o stanie obiektu. Urządzeniem cyfrowym, z którym taki układ współpracuje jest najczęściej komputer.

Sygnały dwuwartościowe wprowadza się przez układy separacji galwanicznej i komutatory półprzewodnikowe. Sygnały krótkotrwałe pochodzące np. od przycisków niestabilnych również przechodzą przez układy separacji galwanicznej i są



5.18. Zestaw urządzeń do wprowadzania danych procesorowych

F — filtr dolnoprzepustowy; *KA* — komutator sygnałów analogowych; *KC* — komutator sygnałów dwuwartościowych; *L* — licznik; *USC* — układy separacji galwanicznej

zapamiętywane w przerzutnikach lub wprowadzane do komputera przez układy przerywania programu. Sygnały tej grupy są wprowadzane w zespołach o liczbie bitów dopasowanej do długości słowa komputera.

Sygnały impulsowe również wprowadza się przez układy separacji galwanicznej. Sygnały są zliczane w licznikach, a stan liczników w określonych odstępach czasu jest wprowadzany do komputera. Sygnały impulsowe mogą być też wprowadzane w trybie przerywania programu pracy komputera.

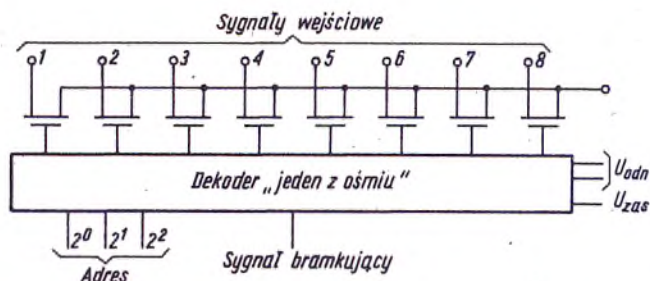
Sygnały analogowe wprowadza się przez filtry dolnoprzepustowe i komutator na wejście przetwornika analogowo-cyfrowego. Stosowane w układach wejściowych przetworniki a/c posiadają kilka zakresów pomiarowych. Właściwy zakres przetwornika jest ustawiany jednocześnie z przyłączeniem przez komutator sygnału wejściowego.

5.19. Jakie elementy stosuje się w komutatorach sygnałów wejściowych urządzeń cyfrowych?

Sposób komutacji zależy od rodzaju sygnałów. Sygnały cyfrowe wprowadza się za pomocą standardowych bramek cyfrowych. Do komutacji sygnałów analogowych stosowane są klucze tranzystorowe z tranzystorami MOS. Obecnie są produkowane zestawy kluczy z tranzystorami MOS w postaci kostek scalonych zawierających od razu dekodery adresu. Schemat ośmiokanałowego komutatora tego typu przedstawiono na rys. 5.19.

Do komutacji słabych sygnałów analogowych (poniżej 100 mV) stosowane są także klucze stykowe (przełączniki), przede wszystkim przełączniki hermetyczne (kontaktrony) o podwyższonej niezawodności.

•5.19• Ośmiobitowy scalony zestaw kluczy



•5.20• Dlaczego sygnały analogowe wprowadzane do urządzeń cyfrowych poddaje się filtracji?

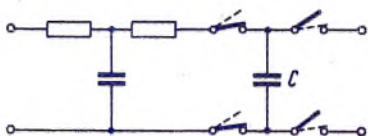
Eliminacja zakłóceń zniekształcających wartości sygnałów analogowych wprowadzanych do cyfrowych urządzeń sterujących stanowi trudny problem techniczny. Przewody doprowadzające sygnały są często długie i prowadzone w obecności silnych pól elektromagnetycznych. Zakłócenia, szczególnie o częstotliwości sieci, mogą więc osiągać stosunkowo duże wartości. Większość przetworników a/c mierzy przy tym wartości chwilowe napięć. Zakłócenia mogą więc znacznie bardziej zniekształcić wynik pomiaru niż w układach analogowych, gdzie stosowana aparatura (np. mierniki wychyłowe) ma własności filtrujące.

Należy przy tym pamiętać, że jednym z powodów stosowania układów cyfrowych jest możliwość zwiększenia dokładności. Żądane dokładności pomiaru wynoszą 0,5 ... 0,1%. Oznacza to, że np. przy wprowadzaniu sygnałów z termoelementów zniekształcenia sygnału wejściowego nie mogą być większe niż kilka do kilkunastu mikrowoltów.

•5.21• Jak można zmniejszyć zakłócenia w torach sygnałów analogowych doprowadzanych do układów cyfrowych?

Trasy kabli doprowadzających sygnały pomiarowe są wybierane bardzo starannie, aby przebiegały możliwie daleko od kabli energetycznych i urządzeń wytwarzających silne pola elektromagnetyczne. Przewody doprowadzające są skręcane i bardzo starannie ekranowane. Często stosuje się połączenia trójprzewodowe — dwa przewody służą do przesyłania sygnału, trzeci do łączenia punktów uziemienia nadajnika i odbiornika sygnału.

Wszystkie te środki zabezpieczające nie eliminują zakłóceń całkowicie. Częstotliwość zakłóceń jest z reguły dużo większa od częstotliwości sygnału użytecznego — większość z nich to zakłócenia o częstotliwości sieci i jej harmoniczne. Stosuje się więc pasywne filtry dolnoprzepustowe. Filtry są niezależne dla każdego z kanałów; zastosowanie filtra wspólnego dla wszystkich kanałów, umieszczonego za komutatorem wprowadziłoby błędy wynikające z powstawania stanów przejściowych w filtrze przy przełączaniu kanałów.



•5.21• Filtr z przełączanym kondensatorem

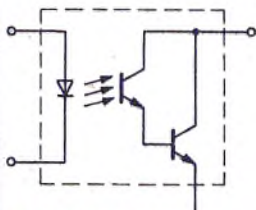
Oprócz zwykłych filtrów RC stosuje się rozwiązania takie, jak na rys. 5.21. W układzie tym kondensator C stanowi jeden z elementów filtru dolnoprzepustowego filtrującego sygnał od czujnika. Gdy czujnik zostanie wybrany przez komutator, wówczas kondensator jest odłączany od filtra i przyłączany do wejścia przetwornika a/c. Dzięki takiemu rozwiązaniu przetwornik jest odłączony od przewodów doprowadzających i zakłócenia mają tylko taki wpływ, że zniekształcają wartość napięcia istniejącego na kondensatorze w chwili odłączania go od źródła sygnału. Stosowane są również rozwiązania, w których kondensator C jest przyłączany do źródła sygnału na okres stanowiący wielokrotność okresu najgroźniejszych zakłóceń (najczęściej są to zakłócenia sieciowe). Wówczas uzyskuje się dla tych zakłóceń tłumienie rzędu 60 dB.

Eliminacja zakłóceń jest tym trudniejsza im niższy jest poziom sygnału, szczególną uwagę trzeba poświęcić sygnałom poniżej 100 mV. Coraz częściej spotykanym rozwiązaniem jest stosowanie indywidualnych wzmacniaczy instalowanych obok źródła sygnału (najczęściej tym źródłem sygnału jest termoelement) i przesyłanie sygnału o napięciu kilku woltów.

•5.22• Po co jest potrzebna i jak jest realizowana separacja galwaniczna pomiędzy obiektem a układem cyfrowym?

Urządzenia cyfrowe, a w szczególności komputery, są budowane z elementów, które mogą ulec uszkodzeniu na skutek wprowadzenia sygnału o niewłaściwej polaryzacji lub wartości (dopuszczalna wartość sygnałów dla najpowszechniej stosowanych elementów wynosi kilka woltów). Sygnały od obiektu do centralnego urządzenia cyfrowego doprowadzane są przewodami o znacznej nieraz długości. Na przewodach tych mogą się pojawić, np. na skutek uszkodzenia urządzeń lub błędu obsługi, napięcia znacznie wyższe od dopuszczalnych. Doprowadzenie takich napięć do urządzenia cyfrowego mogłoby spowodować znaczne, trudne do przewidzenia, uszkodzenia. Przed tego typu uszkodzeniami zabezpieczają układy separacji galwanicznej.

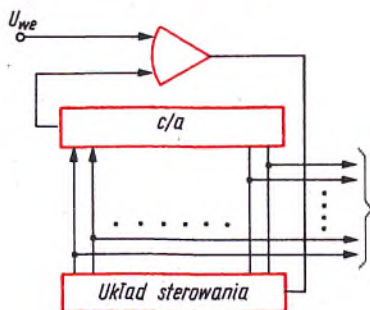
Separację galwaniczną realizuje się obecnie przy użyciu tzw. **optoizolatorów**, tzn. elementów zawierających diodę świecącą i fototranzystor. Doprowadzenie do takiego elementu sygnału „jedyńki” (patrz rys. 5.22) powoduje zaświecenie diody świecącej i pojawienie się „jedyńki” na wyjściu fototranzystora. Optoizolatory zapewniają pełne galwaniczne oddzielenie układu sterującego od jego obwodów wejściowych, dla napięć do 500 V.



•5.22• Schemat optoizolatora

•5.23• Co to jest i jak działa przetwornik analogowo-cyfrowy?

Przetwornikiem analogowo-cyfrowym (przetwornikiem a/c) nazwano układ przetwarzający sygnał analogowy na odpowiadającą jego wartości liczbę. W przetwornikach stosowa-



•5.23• Schemat blokowy przetwornika analogowo-cyfrowego

nych w układach automatyki liczba ta występuje najczęściej w kodzie dwójkowym, a wielkością przetwarzaną jest napięcie prądu stałego. Spośród wielu możliwych rozwiązań przetworników a/c najpowszechniej jest stosowany układ przetwornika z kolejnymi przybliżeniami rys. 5.23.

Napięcie przetwarzane jest podawane na jedno wejście różnicowego wzmacniacza porównującego (komparatora). Na drugie wejście tego wzmacniacza jest podawany sygnał wyjściowy z przetwornika cyfrowo-analogowego (patrz pyt. 5.28) przetwarzającego na napięcie sygnał cyfrowy generowany przez układ sterowania. Cały proces przetwarzania jest realizowany w n krokach — n jest liczbą bitów cyfrowego odpowiednika przetwarzanego napięcia.

W pierwszym kroku układ sterujący wytwarza jedynkę na najbardziej znaczącej pozycji i zera na wszystkich pozostałych pozycjach rejestru. Powoduje to powstanie na wyjściu przetwornika c/a napięcia równego połowie pełnego zakresu. Jeżeli napięcie z przetwornika c/a okaże się mniejsze niż napięcie przetwarzania U_{we} , wówczas jedynka znajdująca się na najbardziej znaczącej pozycji rejestru zostaje na tej pozycji zapamiętana. Jeżeli natomiast napięcie z przetwornika c/a jest większe niż napięcie U_{we} , na najbardziej znaczącej pozycji rejestru zostaje zapamiętane zero.

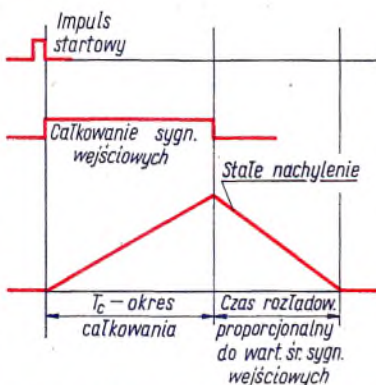
W następnym kroku układ sterujący, pozostawiając na najbardziej znaczącej pozycji rejestru stan taki, jaki został zapamiętany w wyniku wykonania pierwszego kroku, wytwarza sygnał odpowiadający jedynce na następnej z kolei pozycji i sygnały odpowiadające zerom na wszystkich pozostałych pozycjach. Znowu, w wyniku porównania napięcia z przetwornika c/a z napięciem U_{we} , na drugiej pozycji zostanie zapamiętana, zależnie od wyniku tego porównania jedynka lub zero.

W trzecim kroku układ sterujący, pozostawiając na dwóch najbardziej znaczących pozycjach stan taki, jaki został ustalony w wyniku wykonania dwóch pierwszych kroków, generuje jedynkę na trzeciej pozycji i zera na pozostałych. W ten sam sposób są wykonywane wszystkie następne kroki aż do najmniej znaczącej pozycji.

Po wykonaniu ostatniego kroku w rejestrze znajduje się zapamiętany cyfrowy odpowiednik przetwarzanego napięcia. Jest to wielkość wyjściowa przetwornika.

Dokładność przetwarzania w omówionym przetworniku zależy od dokładności pracy przetwornika c/a. W układach automatyki są stosowane najczęściej przetworniki 10 ... 12 bitowe (dokładność 0,1 ... 0,025%); przy obecnym stanie techniki nie są to dokładności trudne do uzyskania.

Przetworniki pracujące tak jak omówiony są obecnie produkowane przez wiele firm jako układy o średnim stopniu scalenia. Do układów takich trzeba jedynie przyłączyć „drabinę” rezystorów (wchodzących w skład przetwornika c/a) oraz komparator.



•5.24A• Przebiegi czasowe w przetworniku analogowo-cyfrowym z podwójnym całkowaniem

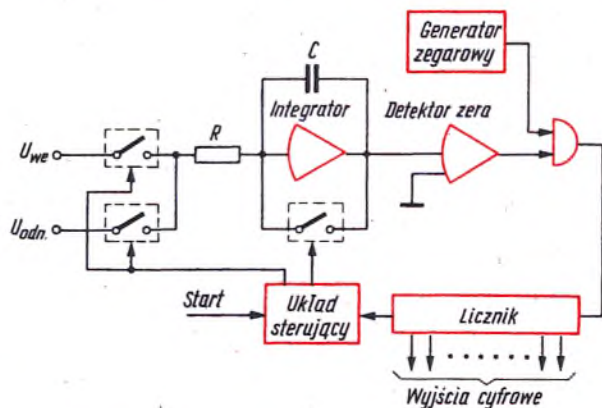
•5.24• Co to jest i jak działa integracyjny przetwornik a/c?

Wielkość cyfrowa (liczba w kodzie dwójkowym) otrzymana na wyjściu przetwornika omówionego w pyt. 5.23 jest proporcjonalna do chwilowej wartości sygnału wejściowego, istniejącej w chwili wykonywania ostatniego kroku przetwarzania. Dlatego właśnie sprawa filtracji sygnałów wejściowych jest tak ważna i jednocześnie bardzo trudna ze względu na żadaną dokładność filtracji — równą dokładności przetwornika. Problem zakłóceń był powodem opracowania integracyjnego przetwornika analogowo-cyfrowego. Na rysunku 5.24A przedstawiono przebiegi czasowe w tzw. przetworniku z dwukrotnym całkowaniem, wyjaśniające zasadę jego działania.

Napięcie wejściowe jest podawane do integratora przez okres czasu T_c . Po tym okresie sygnał wejściowy jest odłączony od integratora, a przyłączone jest napięcie odniesienia o polaryzacji przeciwnej niż sygnał wejściowy. Czas, po którym napięcie wyjściowe integratora stanie się równe zero, jest proporcjonalny do wartości średniej sygnału wejściowego za okres T_c . Wystarczy zmierzyć ten czas w sposób cyfrowy (np. przez zliczanie impulsów o określonej częstotliwości), aby otrzymać wartość średnią sygnału wejściowego w postaci cyfrowej. Jeżeli jako okres T_c przyjmimy wartość okresu sieci, wówczas wynik pomiaru będzie niezależny od zakłóceń o częstotliwości sieci i jej harmonicznych.

Schemat blokowy przetwornika z podwójnym całkowaniem jest przedstawiony na rys. 5.24B. Sygnał wejściowy jest podawany na wejście integratora, którym jest wzmacniacz operacyjny z całkującym sprzężeniem zwrotnym. Przed rozpoczęciem przetwarzania integrator jest zwarty. Impuls startowy powoduje wyzerowanie licznika wyjściowego, podanie sygnału wejściowego do integratora, otwarcie klucza zwierającego integrator oraz otwarcie bramki dla impulsów z generatora zegarowego.

Napięcie na wyjściu integratora narasta przez czas potrzebny do zapełnienia licznika. W chwili przepełnienia licznika wy-



•5.24B• Schemat blokowy przetwornika analogowo-cyfrowego z podwójnym całkowaniem

syłany jest z niego impuls powodujący odłączenie od wejścia integratora sygnału przetwarzanego i przyłączenie do niego napięcia wzorcowego o przeciwnym niż sygnał wejściowy znaku.

Rozpoczyna się rozładowywanie integratora, w czasie którego licznik w dalszym ciągu zlicza impulsy z generatora zegarowego zwiększając swoją zawartość (w chwili przyłączenia do wejścia integratora napięcia wzorcowego licznik był wyzerowany, ponieważ nastąpiło przepełnienie). Gdy napięcie wyjściowe integratora stanie się równe zeru, sygnał wytworzony przez detektor zera zamyka bramkę, przez którą impulsy z generatora zegarowego są wprowadzane do licznika.

Po zakończeniu przetwarzania stan licznika odpowiada wartości napięcia wejściowego; jest on wielkością wyjściową przetwornika.

·5.25· W jaki sposób realizuje się oddziaływanie urządzenia cyfrowego na proces technologiczny?

Oddziaływanie układu sterującego na proces technologiczny jest realizowane za pośrednictwem członów wykonawczych (zasuw, zaworów itp). W większości są to wolno działające elementy o działaniu ciągłym. W układach bezpośredniego sterowania cyfrowego sygnał wyjściowy, który ma regulować przebieg procesu technologicznego, ma postać cyfrową i jest dostępny tylko przez krótki okres czasu określony dla każdego kanału. Ogniwo pośrednie pomiędzy szybko działającym, obsługującym wiele kanałów urządzeniem cyfrowym, a wolno działającymi analogowymi członami wykonawczymi stanowią urządzenia wyjściowe. Powinny one zapewnić:

- Przyporządkowanie sygnału wyjściowego odpowiedniemu kanałowi (komutację);
- Zapamiętanie wartości sygnału;
- Przekształcenie i wzmocnienie mocy sygnału wyjściowego;
- Pozostanie urządzeń wykonawczych w położeniu zajmowanym w chwili awarii bloku liczącego.

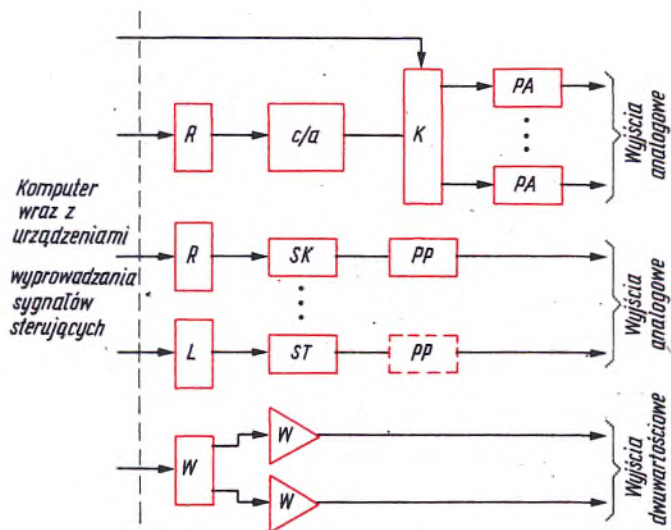
·5.26· Jakie urządzenia są potrzebne aby sygnały cyfrowe można użyć do sterowania obiektów przemysłowych?

Schemat blokowy układu formującego sygnały do sterowania procesem przedstawiono na rys. 5.26. Informacje cyfrowe są zapamiętywane w rejestrach lub licznikach, a następnie przetwarzane do postaci potrzebnej w urządzeniach współpracujących. O wpisaniu informacji przekazywanej z komputera do określonego rejestru lub licznika decyduje adres tego urządzenia przekazywany jednocześnie z informacją.

Sygnały przeznaczone do sterowania cewek zaworów, sprzęgieł, przekładników oraz podobnych urządzeń są zapamiętywane w jednobitowych elementach pamięciowych, np. w przerzutnikach dwustabilnych. Niekiedy stosuje się niezależne źródła zasilania tych przerzutników i inne środki zabezpieczające zapisaną informację przed skasowaniem. Do

5.26 Schemat blokowy układu formującego sygnały do sterowania procesem

R — rejestr; *L* — licznik; *K* — komutator;
PA — pamięć analogowa, *SK* — silnik skokowy;
PP — przetwornik przesunięcie-ciśnienie;
ST — sterowanie czasu włączenia silnika



wzmocniania tych sygnałów używa się kluczowych wzmacniaczy tranzystorowych lub tyrystorowych.

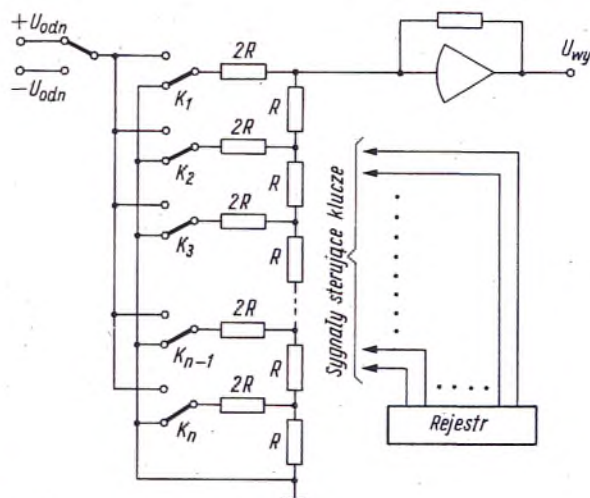
Sygnały przeznaczone do przetworzenia na postać analogową są zapamiętywane, a następnie podawane na wejście przetwornika c/a lub do układów sterowania silników.

5.27 W jaki sposób urządzenie cyfrowe może sterować urządzeniami, w których potrzebne są analogowe sygnały sterujące?

Cyfrowe sygnały wyjściowe przeznaczone do ciągłego sterowania urządzeń wykonawczych (w systemach bezpośredniego sterowania cyfrowego — BSC) oraz do zmian wartości zadanych regulatorów (w systemach sterowania pośredniego) są przetwarzane na napięcia albo na zmiany położenia wału silnika.

Przetwarzania sygnałów cyfrowych na napięcia dokonuje się przy użyciu przetworników c/a. Dalsza obróbka sygnałów zależy od ich przeznaczenia oraz od przyjętego systemu sterowania urządzeń wykonawczych. W szczególności duże znaczenie ma fakt czy w obliczeniach zastosowano algorytm przyrostowy czy pozycyjny. W pierwszym przypadku sygnał wyjściowy określa żadaną zmianę położenia elementu wykonawczego, a w drugim jego położenie żadane. Algorytm przyrostowy jest stosowany częściej niż algorytm pozycyjny ze względu na swoje zalety, do których zaliczamy: zachowywanie, w przypadku awarii bloku liczącego, przez elementy wykonawcze położen zajmowanych ostatnio, możliwość zrezygnowania z sygnału sprzężenia zwrotnego określającego położenie elementu wykonawczego oraz konieczność pamiętania wartości sygnału wyjściowego tylko przez stosunkowo krótki okres czasu.

Przy zamianie sygnałów cyfrowych na położenie wału silnika także najchętniej stosuje się algorytm przyrostowy. Używane



•5.28• Przetwornik cyfrowo-analogowy

są tu silniki skokowe, dla których sygnał cyfrowy określa liczbę „kroków” do wykonania, oraz silniki stałoprędkościowe (np. asynchroniczne), dla których sygnał cyfrowy określa czas włączenia.

•5.28• Jak działa przetwornik cyfrowo-analogowy?

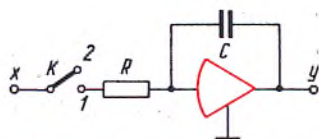
Przetwornik cyfrowo-analogowy (rys. 5.28) umożliwia zamianę wielkości cyfrowych (liczb) na odpowiadające im wielkości analogowe (wartości napięcia).

Przetwarzana liczba podawana w kodzie dwójkowym jest wpisywana do rejestru, którego każda pozycja steruje jednym z kluczy $K_1 \dots K_n$. Wpisanie jedynki do komórki rejestru powoduje zamknięcie odpowiadającego jej klucza. Wypadkowa wartość rezystancji wejściowej wzmacniacza odpowiada kombinacji wpisanych do rejestru zer i jedynek. Przy określonej, stałej wartości napięcia odniesienia napięcie wyjściowe wzmacniacza odpowiada więc liczbie wpisanej do rejestru. Bit znaku przetwarzanej liczby określa polaryzację napięcia wyjściowego. Możliwe do uzyskania dokładności przetwarzania są większe niż wymagana dokładność urządzeń wyjściowych, a uzyskiwane czasy przetwarzania są rzędu mikrosekund (w przypadku stosowania kluczy tranzystorowych).

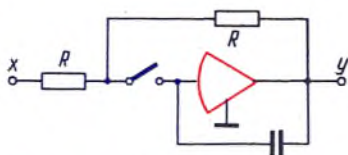
•5.29• Czy jeden przetwornik cyfrowo-analogowy może obsługiwać kilka kanałów?

Tak, jeżeli jego sygnał wyjściowy jest zapamiętywany w analogowych elementach pamięciowych.

Pamięci analogowe stosuje się do zapamiętywania sygnałów sterujących urządzeniami analogowymi. Budowane są one jako układy całkujące (rys. 5.29A). Wzmacniacz operacyjny jest objęty całkującym (pojemnościowym) sprzężeniem zwrotnym. W położeniu 1 klucza K sygnał wejściowy x jest całkowany i napięcie wyjściowe y jest określone zależnością



•5.29A• Schemat pamięci analogowej przyrostowej



•5.29B• Pamięć analogowa pamiętająca wartości ostatniego sygnału

$$y(t_1) = \frac{1}{RC} \int_{t_0}^{t_1} x dt + y(t_0)$$

W położeniu 2 klucza K sygnał wyjściowy y jest utrzymywany na stałym poziomie. Zmiany (zniekształcenia) wartości sygnału y mogą być powodowane:

- Upływnością kondensatora C ;
- Skończoną wartością rezystancji klucza K w stanie rozwarcia;
- Prądem wejściowym wzmacniacza operacyjnego;
- Skończoną wartością rezystancji między wejściem wzmacniacza a masą.

Pamięci analogowe omówionego typu buduje się przy użyciu wzmacniaczy operacyjnych z dodatkowym stopniem wejściowym z tranzystorami MOS. Stopień wejściowy stosuje się w celu uzyskania bardzo dużej rezystancji wejściowej (małego prądu wejściowego). W układach takich uzyskuje się stałość sygnału wyjściowego około 0,3%/h.

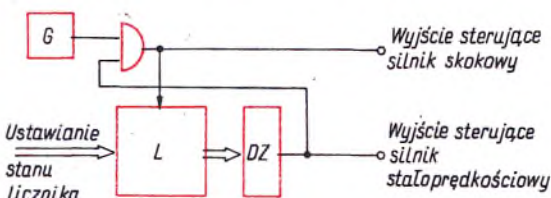
Jak wynika z podanej zależności pomiędzy sygnałem wyjściowym a wejściowym, układ przedstawiony na rys. 5.29A stanowi pamięć, którą można nazwać **pamięcią nagromadzającą**. Jest ona przystosowana do pracy w systemach stosujących przyrostowy algorytm sterowania. W systemach z algorytmem pozycyjnym używa się pamięci analogowej pamiętającej wartość ostatnio wprowadzonego sygnału (rys. 5.29B). Podobnie jak w pamięci nagromadzającej, zapamiętany tu sygnał ulega zniekształceniu i powinien być możliwie często aktualizowany. Częstość powracania do tego samego kanału i uaktualniania zawartości pamięci jest tak doborana, aby zniekształcenia zapamiętanego sygnału mieściły się w dopuszczalnych zakresach błędów przetwarzania sygnału cyfrowego na napięcie.

•5.30• Kiedy i w jaki sposób sygnały cyfrowe zamienia się na przesunięcie?

Sygnały cyfrowe są przetwarzane na przesunięcie wtedy, gdy urządzenie cyfrowe steruje bezpośrednio silnik, gdy współpracuje z regulatorem, w którym wartość zadana jest ustawiana za pomocą silniczka, gdy współpracuje z przetwornikiem przesunięcie-ciśnienie lub w innych podobnych przypadkach. Schemat blokowy typowego układu do zamiany wielkości cyfrowej na przesunięcie przedstawiono na rys. 5.30. Układ może sterować silniki skokowe, może też służyć

•5.30• Schemat blokowy układu sterującego silnik

G — generator; L — licznik; DZ — detektor zera licznika



do włączania innych silników na ściśle określony okres czasu.

Niezależnie od rodzaju silnika żądana wartość zmiany położenia jest wpisywana do licznika impulsów. Stan licznika jest następnie zmniejszany impulsami zegarowymi. Jeżeli zastosowany jest silnik skokowy, wówczas impulsy rozładowujące licznik sterują również silnik. Jeżeli zastosowany jest silnik stałoprędkościowy załączany na określony okres czasu, to czas włączenia jest równy czasowi zerowania licznika. W obu przypadkach wyzerowanie licznika blokuje wejście dla impulsów zegarowych.

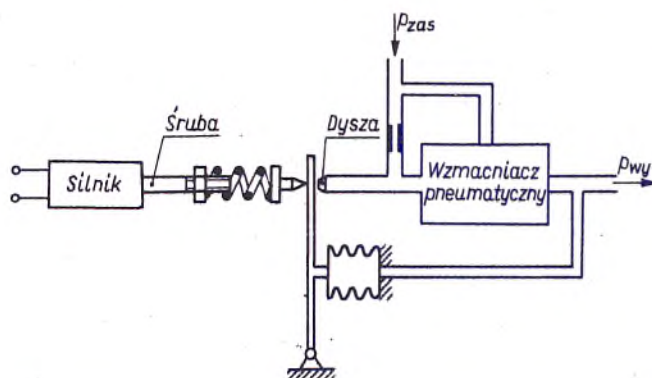
Jako wzmacniacze mocy w tych układach stosowane są układy tranzystorowe i tyrystorowe.

•5.31• W jaki sposób elektryczne urządzenia cyfrowe współpracują z analogową aparaturą pneumatyczną?

Współpraca urządzeń cyfrowych z aparaturą pneumatyczną jest możliwa dzięki stosowaniu odpowiednich przetworników. Schematycznie budowę typowego przetwornika umożliwiającego zamianę sygnału cyfrowego na analogowy sygnał pneumatyczny przedstawiono na rys. 5.31.

Silnik elektryczny sterowany sygnałami wyjściowymi z urządzenia cyfrowego ściska za pomocą przekładni śrubowej sprężynę naciskającą na dźwignię. Z przeciwnej strony niż sprężyna, naciska na dźwignię mieszek sprężysty zasilany wzmocnionym ciśnieniem z układu dysza-przysłona, w którym rolę przysłony spełnia dźwignia. Ciśnienie zasilające mieszek jest ciśnieniem wyjściowym przetwornika, jego wartość ustali się na poziomie potrzebnym do zrównoważenia przez mieszek siły nacisku sprężyny.

Zamianę wielkości cyfrowej na analogową wielkość pneumatyczną uzyskujemy więc za pośrednictwem przetwornika przesunięcie-ciśnienie, w którym przesunięcie jest wymuszane przez silnik sterowany sygnałami cyfrowymi. W przetworniku tym można zastosować silnik skokowy lub silnik o stałej prędkości, włączany na okres czasu proporcjonalny do żądanej zmiany wartości ciśnienia.



•5.31• Przetwornik przesunięcie-ciśnienie

·5.32· Czy niezawodność jest ważnym parametrem urządzeń cyfrowych?

Niezawodność była zawsze uważana za główny parametr techniczny urządzeń cyfrowych stosowanych w automatyce. Wynika to ze skutków ewentualnej awarii cyfrowego urządzenia sterującego. Podstawą działania systemów sterowania cyfrowego jest sterowanie wielu urządzeń technologicznych za pomocą jednego urządzenia cyfrowego. Uzyskuje się to przez podział czasu pracy tego urządzenia między poszczególne kanały. Awaria centralnego urządzenia cyfrowego lub któregośkolwiek z urządzeń zewnętrznych obsługującego wiele kanałów zewnętrznych (np. komutatora) powoduje jednoczesne przerwanie wszystkich kanałów oddziaływania na proces technologiczny i wszystkich kanałów dopływu informacji o przebiegu tego procesu. W takiej sytuacji personel nadzorujący nie jest w stanie przejąć wszystkich funkcji sterowniczych. Jest to istotna różnica w stosunku do klasycznych układów regulacji, w których awaria pojedynczego regulatora wprowadza zakłócenia tylko w jednej pętli i gdzie funkcje tego regulatora mogą być łatwo przejęte przez operatora.

·5.33· W jaki sposób zabezpiecza się układy sterowania przed skutkami awarii komputera sterującego?

Pierwszym warunkiem niedopuszczenia do powstania skutków awarii centralnego układu sterującego jest szybkie wykrywanie stanów awaryjnych. W tym celu stosuje się specjalne programy testowe sprawdzające poprawność działania fragmentów lub całości układu sterowania. Na przykład poprawność działania urządzeń wejściowych (komutatora, wzmacniacza, przetwornika) sprawdza się przez doprowadzenie co jakiś czas (z reguły w stałym cyklu pracy) napięcia o stałej, znanej wartości i porównywanie otrzymanego wyniku z wynikiem prawidłowym. Poprawność pracy urządzeń wyjściowych sprawdza się przez wyprowadzenie w regularnych odstępach czasu określonego sygnału do specjalnego urządzenia alarmowego; zniekształcenie lub nieobecność tego sygnału powoduje ogłoszenie alarmu. Bada się też, czy wyniki obliczeń lub wartości sygnałów wejściowych nie są nieprawdopodobne, tzn. czy ich wartości lub szybkości zmian nie przekraczają granic wynikających z własności obiektu i przyjętych algorytmów obliczeń. Obwody zasilania są objęte oddzielną kontrolą.

Wykrycie awarii powoduje: włączenie alarmu sygnalizującego miejsce i rodzaj awarii, zablokowanie sygnałów wyjściowych urządzenia sterującego oraz włączenie urządzeń rezerwowych, jeżeli takie istnieją. Jako zasadę przyjęto, że sygnały wyjściowe w cyfrowych układach automatyki nie mogą zmieniać swych wartości w wyniku awarii urządzenia sterującego.

Liczba i rodzaj urządzeń rezerwowych zależą od rodzaju

układu sterowania. Problem niezawodności najostrzej występuje w układach bezpośredniego sterowania cyfrowego, tam też układy rezerwowe są najbardziej rozbudowane. W krańcowym przypadku system rezerwowy może obejmować całą część centralną układu sterowania.

·5.34· Jaką rolę w układzie bezpośredniego sterowania cyfrowego spełniają urządzenia rezerwowe?

Urządzenia rezerwowe powinny zabezpieczyć możliwość kontynuowania procesu regulacji w ważniejszych kanałach w przypadku awarii komputera. Niezbędna liczba i rodzaj aparatury rezerwowej zależy od rodzaju procesów i od rozmiarów obiektu technologicznego. Ocenia się, że np. w obiektach chemicznych i rafineryjnych liczba ważnych pętli regulacyjnych, w których musi być prowadzona regulacja niezależnie od awarii komputera wynosi od 10% do 40% wszystkich pętli regulacyjnych.

Jeżeli sterowanie awaryjne ma być realizowane ręcznie, stosowane są stacyjki sterowania ręcznego umożliwiające przejście ze sterowania zaworu sygnałami z komputera na zdalne sterowanie przez operatora. Gdy istnieje konieczność zapewnienia sterowania rezerwowego, w dużej liczbie obwodów regulacyjnych stosuje się analogowe regulatory PI lub PID jako urządzenia awaryjne.

·5.35· Jakie regulatory stosuje się w komputerowych układach sterowania?

W układach BSC oraz w układach sterowania pośredniego stosuje się analogowe regulatory PI oraz PID.

Regulatory pracujące jako urządzenia rezerwowe powinny umożliwiać bezzakłóceniovie przełączanie rodzaju pracy (sterowanie cyfrowe, automatyczno-analogowe, ręczne). Szczególnie ważne jest tu przełączanie ze sterowania cyfrowego na automatyczno-analogowe, ponieważ odbywa się ono w warunkach awarii komputera sterującego, nie ma wtedy możliwości równoważenia sygnałów wyjściowych. Typowym rozwiązaniem problemu bezzakłócenioviego przełączania rodzajów pracy jest zastosowanie na wyjściu regulatora układu pamięci analogowej. Taki właśnie układ zastosowano w omówionym w odpowiedziach na pyt. 4.32 i 4.33 regulatorze ARC-21. Regulator ten może współpracować ze zmodyfikowaną stacją sterowania ręcznego mającą obok pozycji „praca automatyczna” i „praca ręczna” także pozycję BSC, przy której zawartość wyjściowej pamięci analogowej jest zmieniana sygnałami z komputera.

W regulatorach pracujących jako urządzenia rezerwowe oraz w regulatorach pracujących w układach sterowania pośredniego istnieje potrzeba wprowadzania zmian wartości zadanej za pomocą sygnałów z komputera. Najczęściej zmiany wartości zadanej regulatorów uzyskuje się za pomocą silniczków przedstawiających potencjometri wartości zadanej. Rzadziej są stosowane układy pamięci analogowych z układami anty-

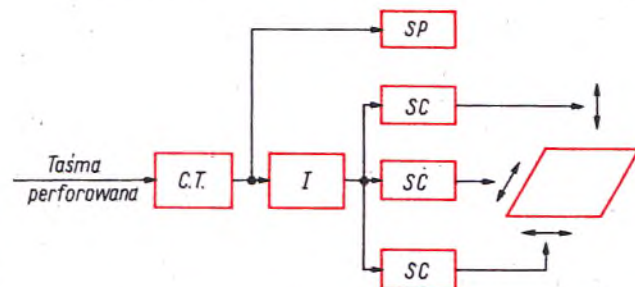
dryftowymi. Jak widać, zmiany wartości zadanej regulatorów uzyskuje się tymi samymi środkami technicznymi co zmiany położenia zaworów.

W niektórych przypadkach nie zmienia się wartości zadanej sygnałami z komputera, a zamiast tego w regulatorze jest ustawiona stała „bezpieczna” wartość zadana. W przypadku awarii komputera sterującego regulacja jest dalej prowadzona wg tej wartości zadanej. Takie rozwiązanie jest szczególnie przydatne w układach, w których proces jest prowadzony przez komputer blisko ograniczeń awaryjnych i bez nadzoru komputera mogłoby nastąpić przekroczenie tych ograniczeń. Przełączanie do „bezpiecznej wartości zadanej” stosuje się również przy sterowaniu pośrednim w przypadku awarii komputera obliczającego nastawy regulatorów.

•5.36• Jak są zbudowane i gdzie się stosuje układy programowego sterowania położenia?

Schemat blokowy układu programowego sterowania frezarki jest przedstawiony na rys. 5.36. Program obróbki zapisany na papierowej taśmie perforowanej, po odczytaniu i zdekodowaniu jest wprowadzany do interpolatora. Interpolator, otrzymując jako dane wejściowe współrzędne (lub zmiany współrzędnych) kolejnych punktów toru, po którym powinno poruszać się narzędzie (frez), wytwarza na swoich wyjściach ciągi impulsów sterujących cyfrowe serwomechanizmy wykonawcze. Przebiegi wyjściowe z interpolatora sterują serwomechanizmy w taki sposób, że narzędzie porusza się między współrzędnymi początkowymi i końcowymi po linii prostej (jeżeli interpolator jest liniowy) lub po krzywej drugiego stopnia. Niektóre rozkazy są kierowane do układów sterujących czynności dodatkowe, takimi jak zmiana narzędzia, włączenie lub wyłączenie czynnika chłodzącego itp.

Układy programowego sterowania położenia najpierw opracowano do sterowania obrabiarek, dlatego jako przykład podano układ sterowania frezarki. Obecnie zakres ich zastosowań jest znacznie szerszy, są stosowane do cięcia blach, w automatach spawalniczych, do cięcia materiałów włókienniczych, w rejestratorach cyfrowych, w układach do automatycznego montażu (np. central telefonicznych) i w wielu innych zastosowaniach.



•5.36• Układ programowego sterowania frezarki

CT — czytnik taśmy papierowej; I — interpolator; SC — serwomechanizm cyfrowy; SP — sterowanie czynności pomocniczych

·5.37· Co to jest serwomechanizm cyfrowy?

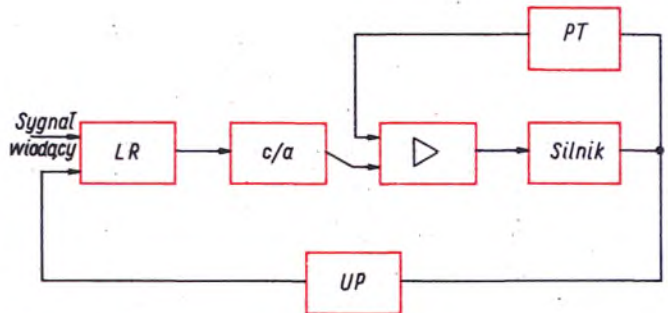
Serwomechanizmem cyfrowym nazywamy układ regulacji położenia, w którym zarówno **sygnał wiodący** jak i **sygnał położeniowego sprzężenia zwrotnego** mają postać cyfrową. Serwomechanizmy cyfrowe stosuje się tam, gdzie sygnał wiodący jest przygotowany przez urządzenia cyfrowe oraz tam, gdzie wymagana jest duża względna dokładność ustawiania mechanizmów (duża dokładność przy dużym zakresie przesunięć).

Sygnał wiodący w postaci ciągu impulsów jest podawany na wejście licznika rewersyjnego (rys. 5.37A). Do tego samego licznika wprowadza się impulsy sprzężenia zwrotnego wytwarzane przez układ pomiaru położenia. Zarówno impulsy zadające jak też impulsy pomiarowe mogą mieć znak „+” lub „-” i zależnie od tego znaku są dodawane lub odejmowane od zawartości licznika. Przy ustalonym kierunku ruchu impulsy pomiarowe mają znak przeciwny niż impulsy zadające. Zawartość licznika rewersyjnego jest w każdej chwili równa algebraicznej różnicy impulsów zadających i pomiarowych (z uwzględnieniem przypisanych im znaków), czyli uchybowi serwomechanizmu.

Zawartość licznika rewersyjnego jest zamieniana przez przetwornik c/a na wartość napięcia i po wzmocnieniu steruje silnik wykonawczy. Sygnał sprzężenia tachometrycznego wytwarzany np. przez prądnicę tachometryczną jest na wejściu wzmacniacza sumowany z sygnałem wyjściowym przetwornika c/a.

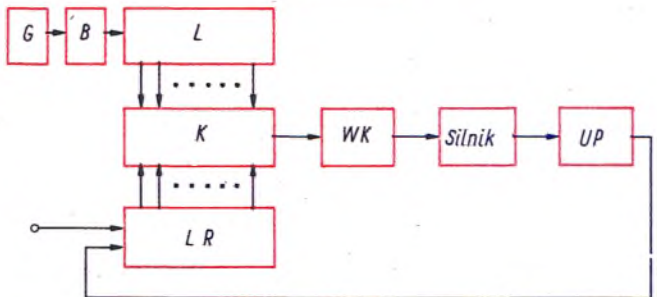
·5.37A· Schemat blokowy serwomechanizmu cyfrowego

LR — licznik rewersyjny; UP — układ pomiaru położenia; PT — prądnica tachometryczna



·5.37B· Schemat blokowy serwomechanizmu cyfrowego z dynamicznym przetwarzaniem wartości uchybu na sygnał sterujący silnik

LR — licznik rewersyjny; L — licznik pomocniczy; K — komparator; G — generator; B — zespół bramkujący; WK — wzmacniacz kluczkowany; UP — układ pomiaru położenia



Na rysunku 5.37B przedstawiono schemat blokowy innego rozwiązania serwomechanizmu cyfrowego. Tak jak w poprzednim rozwiązaniu zawartość licznika rewersyjnego jest równa uchybowi serwomechanizmu, inne jest jednak przebieganie od zawartości licznika do sygnału sterującego silnikiem. Zawartość licznika rewersyjnego jest porównywana w komparatorze z zawartością licznika pomocniczego. Na wyjściu komparatora otrzymywany jest sygnał 1, gdy zawartość licznika rewersyjnego jest większa lub równa zawartości licznika pomocniczego oraz sygnał 0, gdy zawartość licznika rewersyjnego jest mniejsza od zawartości licznika pomocniczego. Do licznika pomocniczego są wprowadzane impulsy z generatora więc jego zawartość zmienia się w sposób ciągły od zera do wartości maksymalnej, osiągając po przepełnieniu stan zero. Proces ten powtarza się cyklicznie z częstotliwością równą ilorazowi częstotliwości impulsów z generatora i pojemności licznika pomocniczego. Cyklicznie powtarzają się również stany, w których zawartość licznika pomocniczego jest mniejsza oraz stany w których, ta zawartość jest większa od zawartości licznika rewersyjnego. W efekcie, na wyjściu komparatora otrzymujemy cyklicznie stan 1 i stan 0, czyli przebieg prostokątny. Częstotliwość tego przebiegu jest równa ilorazowi częstotliwości impulsów z generatora i pojemności licznika pomocniczego, a wypełnienie jest proporcjonalne do zawartości licznika rewersyjnego czyli od wartości uchybu położenia serwomechanizmu.

Sygnał wyjściowy komparatora po wzmocnieniu steruje silnik wykonawczy serwomechanizmu. Sprężenie prędkościowe może być w omawianym układzie zrealizowane w sposób cyfrowy przez zależne od częstotliwości impulsów pomiarowych, oddziaływanie na impulsy wejściowe do licznika pomocniczego. Sygnał sterujący w postaci przebiegu prostokątnego o zmiennym wypełnieniu umożliwia stosowanie kluczowych wzmacniaczy tranzystorowych i wzmacniaczy tyrystorowych. Można dzięki temu stosować wzmacniacze półprzewodnikowe w serwomechanizmach z silnikami wykonawczymi o mocach nawet kilku kilowatów.

Rozdział 6

Optymalizacja i jej zastosowanie w automatyce

·6.1· Co to znaczy „optymalnie”?

Optymalnie to znaczy najlepiej. Jest to pojęcie ostatnio bardzo często używane — w zasadzie we wszystkich dziedzinach działalności człowieka mówi się o optymalizacji, decyzjach optymalnych, optymalnym zarządzaniu itp. Jednocześnie pojęcie to jest źródłem nieporozumień i jest używane niekiedy w sposób sugerujący niezrozumienie istoty problemu optymalizacji. I tak np. spotyka się czasem w prasie lub w rozmowach stwierdzenie, że coś „jest mniej lub bardziej optymalne”. Oczywiście jest to stwierdzenie bez sensu. Jeżeli jakaś decyzja jest optymalna, to to stwierdzenie jednoznacznie ją charakteryzuje — jest to decyzja najlepsza z możliwych. Fakt ten można wykorzystać do zdefiniowania **decyzji optymalnej** — jest to taka decyzja, od której każde odstępstwo jest niekorzystne.

·6.2· Co to znaczy, że decyzja jest najlepsza z możliwych?

Aby odpowiedzieć na to pytanie posłużymy się przykładem. Załóżmy, że mamy przebyć drogę z miasta *A* do miasta *B*. Można to zadanie wykonać wieloma sposobami — np. pojechać pociągiem, autobusem, pójść pieszo. Wybierając jedną z możliwości (a więc podejmując decyzję) narażamy się na określone koszty podróży. Tak więc każdej decyzji można przyporządkować pewną liczbę, określającą w tym przypadku koszt biletu. Nie jest to oczywiście jedyna wartość, jaką można przyporządkować decyzji. Inną wartością może być np. czas podróży, wygoda podróżowania itp. Potrafimy więc sformułować pewną **ilościową** ocenę podejmowanych rozwiązań. Ocenę tę nazywamy **wskaźnikiem jakości decyzji**. Mając taką ocenę ilościową możemy porównywać rozwiązania. Tak więc decyzja najlepsza — to decyzja najlepsza pod względem danego wskaźnika jakości. Podkreślić należy, że wybór wskaźnika jakości jest arbitralny i decyzja optymalna w sensie jednego wskaźnika może nie być optymalna w sensie innego. Rozważmy dla ustalenia uwagi sformułowany wyżej przykład. Jeżeli wskaźnikiem jakości będzie czas podróży, to decyzją optymalną będzie wybór samolotu jako środka

podróży; jeżeli wskaźnikiem jakości będzie koszt — to rozwiązaniem optymalnym będzie podróż pieszo. Tak więc wskaźnik jakości jest podstawowym składnikiem zadania optymalizacji i jego wybór rzutuje na końcowe rozwiązanie problemu.

•6.3. Jak formuluje się wskaźnik jakości?

Odpowiedź na to pytanie nie jest prosta. Ze względu na wagę problemu sformułowaniu wskaźnika jakości należy poświęcić wiele uwagi. Często sformułowanie wskaźnika jest znacznie trudniejsze, niż rozwiązanie zadania optymalizacji i wymaga współdziałania specjalistów z wielu dziedzin, np. inżynierów i ekonomistów.

Sformułowanie wskaźnika nie przedstawia większych kłopotów w przypadku, gdy ma on sens ekonomiczny, np. jest to koszt związany z realizacją decyzji. W przykładzie z poprzedniego punktu jest to po prostu koszt podróży. Zadanie komplikuje się znacznie, gdy podejmując decyzję trzeba uwzględniać wiele czynników nieporównywalnych ze sobą, np. koszt i czas podróży. Pierwszy z nich wyrażony jest w jednostkach pieniężnych, a drugi w godzinach. Możliwe są w zasadzie dwa sposoby rozwiązania problemu. Pierwszy z nich, tzw. **metoda cen**, polega na sprowadzeniu wszystkich składników oceny decyzji do „wspólnego mianownika”, tzn. wyrażenie ich w tych samych jednostkach. Tak więc według powiedzenia, że „czas to pieniądz” można przypisać czasowi podróży pewną cenę. Cena ta może określać np. straty wynikłe z niedostarczenia na czas części zamiennej przez pracownika serwisu fabrycznego. W analogiczny sposób można przypisać pewną cenę komfortowi podróży — pracownik podróżując w złych warunkach będzie musiał odpoczywać pewien czas zanim zabierze się do precyzyjnej naprawy, a więc zwiększy się czas przestoju urządzenia.

Ponieważ wszystkie składowe wskaźnika jakości wyraziliśmy w jednostkach pieniężnych, można napisać wyrażenie określające jakość podejmowanej decyzji

$$J = k_p + c_t \cdot t + c_k$$

gdzie: k_p oznacza koszt podróży; t — czas podróży; c_t — koszt związany z czasem podróży; c_k — koszt związany z komfortem podróży, przy czym ma ona wartość największą dla podróży piechotą, a najmniejszą dla podróży samolotem.

Tak więc potrafimy już porównywać podejmowane decyzje — z dwóch decyzji ta jest lepsza, która zapewnia mniejszą wartość wskaźnika J .

Drugi sposób formułowania wskaźnika jakości polega na wprowadzeniu **wektorowych wskaźników jakości**.

•6.4. Co to jest wektorowy wskaźnik jakości i w jaki sposób służy do porównywania decyzji?

Powróćmy jeszcze raz do przykładu z pyt. 6.2. Jak wspomniano, jakość podejmowanej decyzji można scharakteryzować wprowadzając trzy parametry: koszt, czas i komfort podróży. Ustawiając te trzy wielkości w tablicę

$$J = \begin{bmatrix} k_p \\ t \\ k \end{bmatrix}$$

otrzymamy wektor odpowiadający określonej decyzji tzw. **wektor jakości** lub **wektorowy wskaźnik jakości**. Pojawia się problem: czy wektor ten może służyć do porównywania decyzji, a jeżeli tak, to w jaki sposób?

Założmy, że mamy porównać dwie decyzje d_1 i d_2 scharakteryzowane dwoma wektorowymi wskaźnikami jakości

$$J(d_1) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad J(d_2) = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Otóż mówimy, że decyzja d_1 jest lepsza od decyzji d_2 , jeżeli zachodzą nierówności

$$a_i \leq b_i \quad \text{dla } i = 1, 2, \dots, n$$

Oznacza to, że wszystkie składowe wektora jakości odpowiadające decyzji lepszej mają być nie większe od odpowiednich składowych wektora jakości decyzji gorszej. Fakt ten zapisujemy następująco; $J(d_1) \leq J(d_2)$.

Oczywiście nie wszystkie decyzje można porównywać w sensie takiego wskaźnika jakości.

•6.5• Czy sformułowanie wskaźnika jakości wystarcza do rozwiązania zadania optymalizacji?

Sformułowanie wskaźnika jakości jest zaledwie jednym z etapów formułowania zadania optymalizacji. Do pełnego postawienia problemu należy określenie **zbioru decyzji możliwych i zbioru decyzji dopuszczalnych**. Definicja zbioru decyzji możliwych wynika bezpośrednio z nazwy — jest to zbiór wszystkich możliwych decyzji, które moglibyśmy podjąć rozwiązując dany problem.

Zbiór decyzji dopuszczalnych jest to podzbiór (pewna część) zbioru decyzji możliwych, określony przez ograniczenia obowiązujące przy rozwiązywaniu danego problemu. Konieczność wprowadzenia ograniczeń wynika z faktu, że nie wszystkie decyzje jesteśmy w stanie zrealizować. Może to wynikać ze względów konstrukcyjnych, finansowych itp.

W rozpatrywanym poprzednio przykładzie decyzja polegała na wyborze odpowiedniego środka lokomocji. Tak więc zbiór decyzji możliwych jest to zbiór wszystkich środków lokomocji łączących miasta A i B . Ponieważ środków tych jest skończona liczba, to można je ponumerować liczbami

Nr	Środek lokomocji	Czas [h]	Cena [zł]	Niewygoda
1	Samolot	1	500	1
2	Pociąg	6	200	5
3	Autobus	7	300	10
4	Pieszko	100	200	500

naturalnymi. Tak więc zbiór decyzji możliwych odpowiada zbiorowi

$\{1, 2, 3, 4\}$

Założmy teraz, że podróżny rozporządza kwotą 400 zł, przy czym nie może podróżować dłużej niż 10 godzin. Oczywiście wszystkie decyzje nie spełniające tych warunków należy z góry odrzucić; decyzje spełniające warunki tworzą zbiór decyzji dopuszczalnych (tabl. 6.5). Składa się on tylko z dwóch elementów $\{2, 3\}$.

•6.6• Co można powiedzieć o zbiorze decyzji możliwych?

Postać elementów zbioru decyzji możliwych wynika bezpośrednio z postawienia zadania. Omówimy krótko podstawowe zadania optymalizacji i odpowiadające im zbiory decyzji możliwych.

— Zbiór liczb całkowitych

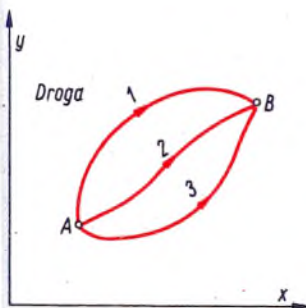
Taki przypadek występuje wówczas, gdy decyzje możliwe można ponumerować kolejno (jak w przykładzie z pyt. 6.5) lub gdy stoimy przed zadaniem rozdziału pewnych zasobów, których elementy są niepodzielne. Typowym przykładem jest tu problem przydziału robotników do wykonania jakiegoś zadania. Tego typu zadania optymalizacji nazywamy zadaniami optymalizacji **całkowitoliczbowymi**, lub **dyskretnymi**.

— Zbiór liczb rzeczywistych

Jest to przypadek najczęściej spotykany w praktyce. Przykładem może być zadanie optymalnego rozdziału strumienia reagentów między dwa reaktory chemiczne, wybór optymalnych inwestycji w jakiejś gałęzi przemysłu itp.

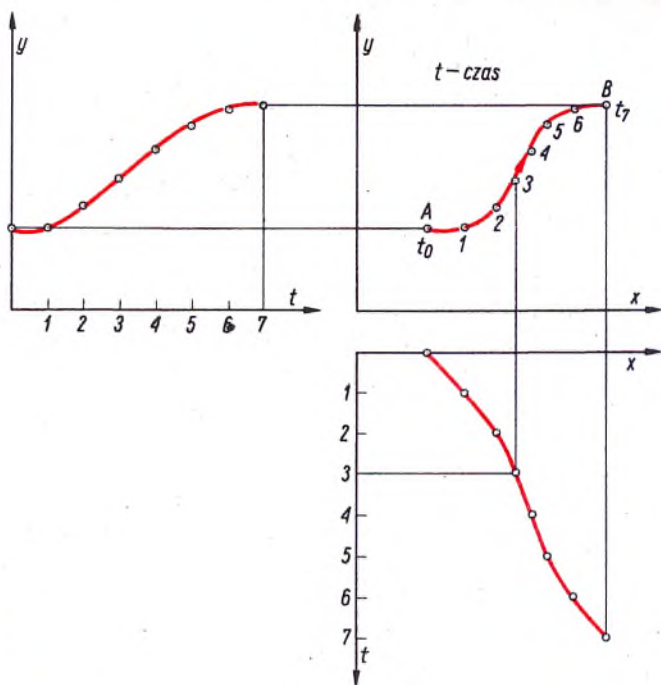
— Zbiór funkcji o pewnych właściwościach

W tym przypadku zagadnienie jest bardziej skomplikowane; wyjaśnimy je na przykładzie. Założmy, że mamy przebyć pieszo drogę z punktu A do punktu B . Można to zrobić wieloma sposobami; każdą wybraną drogę przedstawiamy w postaci pewnej krzywej ciągłej na płaszczyźnie, o początku w punkcie A i końcu w punkcie B (rys. 6.6A). Wybierając określoną drogę (podejmując decyzję) można wyznaczyć pewną wartość charakteryzującą ten wybór, np. czas podróży lub bardziej złożony wskaźnik (np. taki jak w punkcie 6.3). Tak więc każdej krzywej można przyporządkować pewną wartość. Tego typu przyporządkowanie nazywa się **funkcją**.



•6.6A• Ilustracja zadania wariacyjnego — punkty A i B można połączyć wieloma krzywymi; każda z nich może charakteryzować się inną wartością wskaźnika jakości

•6.6B• Przykład sprowadzenia zadania poszukiwania krzywej łączącej punkty A i B do zadania poszukiwania dwóch funkcji czasu



cjonalem. Równanie drogi na płaszczyźnie o współrzędnych prostokątnych x i y można zapisać

$$x = f_1(t)$$

$$y = f_2(t)$$

przy czym zapis ten należy rozumieć następująco (rys. 6.6B): x i y są współrzędnymi punktu, w którym znajduje się podróżny w chwili t (t oznacza czas), f_1 i f_2 są to pewne reguły przyporządkowujące liczbie (czas) liczby (współrzędne punktów), a więc są to funkcje. Wybór drogi sprowadza się zatem do doboru funkcji $f_1(t)$ i $f_2(t)$; są to elementy zbioru decyzji możliwych.

Tego typu problemy optymalizacji nazywamy problemami **wariacyjnymi lub dynamicznymi**.

•6.7• W jaki sposób można określić zbiór decyzji dopuszczalnych?

Sposób zadawania zbioru decyzji dopuszczalnych zależy od natury zbioru decyzji możliwych. Najprostszy i najczęściej spotykany sposób polega na podaniu układu równań i nierówności, jakie muszą spełniać decyzje, np.

$$g_1(d_1, d_2, d_3) = 0$$

$$g_2(d_1, d_2, d_3) \leq 0 \text{ itd.}$$

przy czym d_1, d_2, d_3 należą do zbioru decyzji możliwych D . W przypadku, gdy mamy do czynienia z problemem wariacyjnym, elementami zbioru D są funkcje i zależności okre-

śląjące zbiór decyzji dopuszczalnych mogą być bardziej złożone, np. mogą zawierać pochodne i całki funkcji. Zwróćmy uwagę na ograniczenia o podanej postaci. Tego typu ograniczenia pozwalają na ogół na wyeliminowanie jednej ze zmiennych, np. zmienną d_1 można wyznaczyć za pomocą zmiennych d_2 i d_3 . W najprostszym przypadku równanie może mieć postać

$$d_1 + d_2 + d_3 = 0$$

Wynika stąd, że

$$d_1 = -d_2 - d_3$$

Obecność takich ograniczeń powoduje, że znajomość dwóch zmiennych jednoznacznie wyznacza trzecią. Z tego względu ograniczenie takie nazywamy **więzami**.

•6.8. Jakie można podać przykłady zbiorów decyzji dopuszczalnych?

a) Zbiór decyzji możliwych jest zbiorem liczb rzeczywistych; zakładamy, że mamy do czynienia z zadaniem o dwóch zmiennych decyzyjnych.

W najprostszym przypadku ograniczenia mogą mieć postać

$$d_{1min} \leq d_1 \leq d_{1max}$$

$$d_{2min} \leq d_2 \leq d_{2max}$$

Parę decyzji (d_1, d_2) można interpretować jako punkt na płaszczyźnie, zbiór decyzji dopuszczalnych jest prostokątem (rys. 6.8A).

Ograniczenia mogą mieć bardziej złożoną postać

$$\alpha d_1 + \beta d_2 \leq \gamma$$

gdzie α, β, γ — stałe.

Dla znaku równości otrzymujemy

$$d_2 = -\frac{\alpha}{\beta} d_1 + \frac{\gamma}{\beta}$$

Jest to równanie prostej o współczynniku kierunkowym $-\frac{\alpha}{\beta}$, i przechodzącej przez punkt $d_1 = 0, d_2 = \frac{\gamma}{\beta}$. Jeżeli ustalimy wartość d_1 (pierwsza decyzja) i zmieniamy decyzję d_2 , to dla punktów (d_1, d_2) leżących nad prostą P jest spełniona nierówność

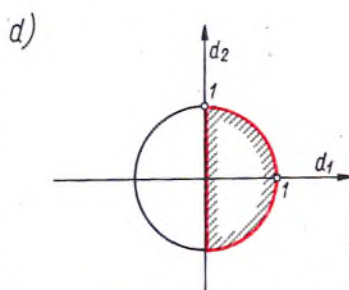
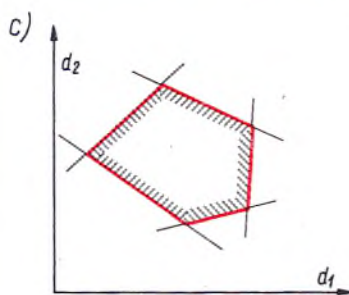
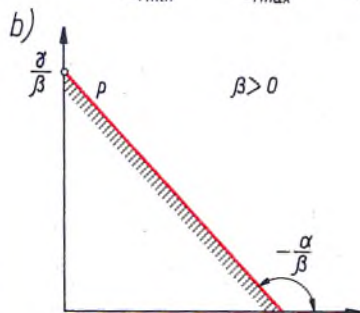
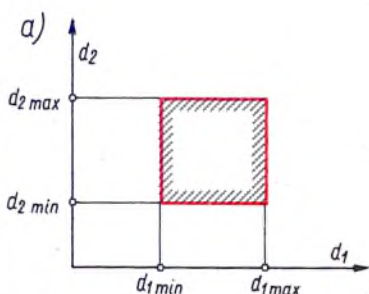
$$d_2 > -\frac{\alpha}{\beta} d_1 + \frac{\gamma}{\beta}$$

Jeżeli β jest dodatnie, to nierówność ta jest równoważna nierówności

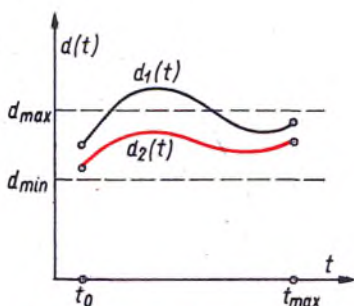
$$\alpha d_1 + \beta d_2 > \gamma$$

W przeciwnym przypadku

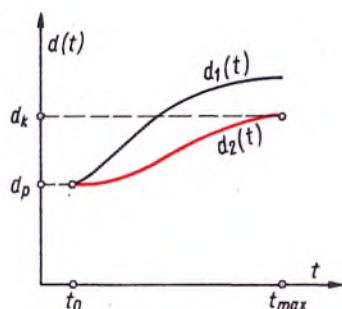
$$\alpha d_1 + \beta d_2 < \gamma$$



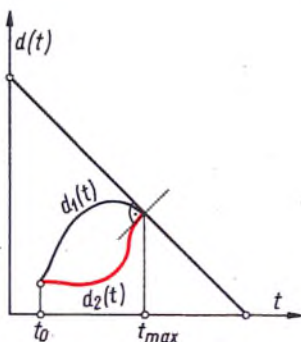
•6.8A. Przykłady zbiorów decyzji dopuszczalnych



•6.8B• Ograniczenia chwilowe wartości funkcji: d_2 — decyzja dopuszczalna; d_1 — decyzja niedopuszczalna (wartości funkcji $d_1(t)$ wykraczają poza dozwolony przedział)



•6.8C• Ograniczenie wartości końcowej i początkowej funkcji. Funkcja $d_2(t)$ jest dopuszczalna, gdyż przyjmuje na końcu i na początku zadane wartości, $d_1(t)$ jest niedopuszczalna



•6.8D• Warunki transversalności — krzywa będąca wykresem funkcji $d(t)$ musi być prostopadła do pewnej prostej $d_2(t)$ — funkcja dopuszczalna; $d_1(t)$ — niedopuszczalna

Tak więc zbiór decyzji dopuszczalnych jest w tym przypadku półprzestrzenią ograniczoną przez prostą P o równaniu $\alpha d_1 + \beta d_2 = \gamma$

przy czym znak β decyduje o tym, po której stronie prostej znajdują się decyzje dopuszczalne.

Jeżeli zbiór decyzji dopuszczalnych jest określony większą liczbą nierówności, to decyzje dopuszczalne powinny spełniać jednocześnie wszystkie nierówności określające ten zbiór. Na rysunku 6.8A przedstawiono zbiór dopuszczalnych określony przez układ 5 nierówności.

Oczywiście w realnych zadaniach zbiór decyzji dopuszczalnych może być bardziej skomplikowany. Rozpatrzmy np. układ nierówności

$$d_1^2 + d_2^2 \leq 1$$

$$d_1 \geq 0$$

Zbiór dopuszczalny przedstawiono na rys. 6.8A(d) — jest to połowa zbioru ograniczonego okręgiem o promieniu jednostkowym i środku w początku układu współrzędnych.

b) Zbiór decyzji możliwych jest zbiorem funkcji (zadanie wariacyjne).

Decyzjami są w tym przypadku funkcje o określonych właściwościach, np. funkcje ciągłe. Ograniczenia mogą być dwóch typów

— ograniczenia chwilowe (punktowe)

— ograniczenia funkcjonalne (globalne).

Ograniczenia punktowe określają właściwości funkcji dla wszystkich (lub niektórych) wartości zmiennej niezależnej. Mogą one mieć np. postać

$$d_{\min} \leq d(t) \leq d_{\max} \quad 0 \leq t \leq t_{\max}$$

Ograniczenie takie mówi, że funkcja (decyzja) jest dopuszczalna, jeżeli w każdej chwili czasu wartości tej funkcji nie wykraczają poza zadany przedział (rys. 6.8B). Ograniczenia funkcjonalne dotyczą właściwości funkcji w całym przedziale zmienności zmiennej niezależnej. Inne, często spotykane ograniczenia dotyczą właściwości funkcji $d(t)$ na końcach przedziału $[t_0, t_{\max}]$. W najprostszym przypadku żąda się, aby funkcja $d(t)$ przyjmowała określone wartości dla t_0 i $t_{\max}$ (rys. 6.8C). W bardziej skomplikowanych przypadkach żądać możemy aby funkcja $y = d(t)$ była w chwili $t_{\max}$ prostopadła do pewnej prostej (rys. 6.8D), np. prostej o równaniu

$$y = -t + 1 \quad (6.8.1)$$

Analityczny zapis takiego warunku wymaga znajomości pewnych faktów z geometrii analitycznej.

Jak wiadomo, dwie proste o równaniach

$$a_1 x_1 + b_1 x_2 = c_1$$

$$a_2 x_1 + b_2 x_2 = c_2$$

są prostopadłe, jeżeli współczynniki spełniają warunek

$$a_1 a_2 + b_1 b_2 = 0$$

Jeżeli chcemy, aby funkcja $d(t)$ była prostopadła do prostej określonej równaniem (6.8.1) w chwili t_{max} , to styczna do tej funkcji w chwili t_{max} powinna być prostopadła do prostej. Równanie stycznej można podać znając pochodną funkcji $d(t)$. Ma ono postać

$$y = d'(t)t + c$$

gdzie c — pewna stała.

Warunek prostopadłości można teraz zapisać

$$1 + d'(t_{max})t_{max} = 0$$

Tego typu warunki nazywamy **warunkami transversalności**. Jak wiadomo z poprzednich rozważań, każdej funkcji przypisać można pewną liczbę, a przyporządkowanie takie nazywamy funkcjonałem. Możemy więc napisać

$$F(d(t)) = a$$

gdzie a jest liczbą przyporządkowaną decyzji d .

Ograniczenie funkcjonałowe ma na przykład postać

$$F(d(t)) \leq F_{max}$$

W najprostszym przypadku funkcjonał może być po prostu całką funkcji $d(t)$; ograniczenie funkcjonałowe będzie następujące

$$\int_0^T d(t) dt \leq F_{max}$$

Oznacza to, że spośród wszystkich możliwych funkcji możemy rozważać tylko takie, dla których pole figury zawartej między wykresem funkcji a osią t jest nie większe niż F_{max} .

•6.9• Co to jest zadanie programowania matematycznego?

Zadaniem programowania matematycznego nazywamy zadanie następujące:

- dany jest zbiór decyzji możliwych,
- dany jest zbiór decyzji dopuszczalnych,
- dana jest pewna funkcja określona na zbiorze decyzji możliwych.

Należy znaleźć taki element zbioru decyzji dopuszczalnych dla którego funkcja osiąga minimum (lub maksimum). Element taki nazywamy **rozwiązaniem zadania programowania matematycznego**.

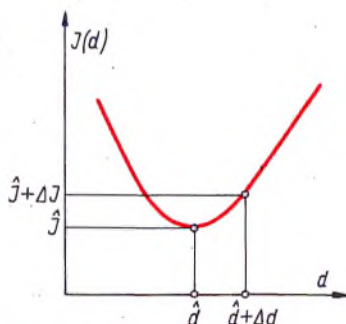
•6.10• Czy istnieje klasyfikacja zadań programowania matematycznego?

Klasyfikację zadań programowania matematycznego przeprowadza się biorąc pod uwagę dwa czynniki:

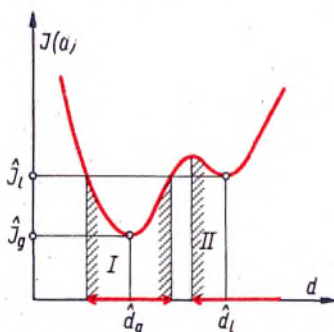
- Właściwości wskaźnika jakości (minimalizowanej funkcji);
- Właściwości zbioru decyzji dopuszczalnych.

Podstawowe klasy zadań programowania matematycznego są następujące:

1. Programowanie liniowe — funkcja liniowa, ograniczenia liniowe;
2. Programowanie liniowe całkowitoliczbowe — jak wyżej, przy dodatkowym założeniu, że decyzje dopuszczalne mogą przyjmować tylko wartości całkowite;
3. Programowanie boolowskie (logiczne) — jak w p. 2, przy dodatkowym założeniu, że decyzje dopuszczalne mogą przyjmować tylko dwie wartości — 0 i 1;
4. Programowanie nieliniowe — co najmniej jedna z funkcji występująca w zadaniu nie jest liniowa;
5. Dynamiczne zadanie programowania matematycznego — zbiór decyzji dopuszczalnych jest zbiorem funkcji, a wskaźnik jakości — funkcjonalem.



•6.11A• Ilustracja pojęcia decyzji optymalnej. Decyzja $\hat{d}$ jest optymalna, gdyż każda inna decyzja zapewnia wzrost wartości funkcji



•6.11B• Minimum lokalne a globalne. W punkcie $\hat{d}_l$ funkcja ma minimum lokalne, gdyż istnieje takie otoczenie tego punktu (obszar II), że w tym obszarze wartości funkcji są większe, niż w punkcie $\hat{d}_l$. W obszarze I wartości funkcji są mniejsze niż w punkcie $\hat{d}_l$. W punkcie $\hat{d}_g$ funkcja ma minimum globalne, gdyż dla dowolnego, innego punktu wartości funkcji są większe, niż w punkcie $\hat{d}_g$.

•6.11• Co należy rozumieć przez rozwiązanie zadania optymalizacji?

Decyzję $\hat{d}$ nazywamy rozwiązaniem zadania optymalizacji (lub decyzją optymalną), jeżeli spełnia ona następujące warunki:

- Należy do zbioru decyzji dopuszczalnych;
- Każda inna decyzja z tego zbioru zapewnia nie lepszą wartość w wskaźniku jakości (rys. 6.11A).
Mówimy wówczas, że decyzja ta zapewnia optimum globalne wskaźnika jakości.

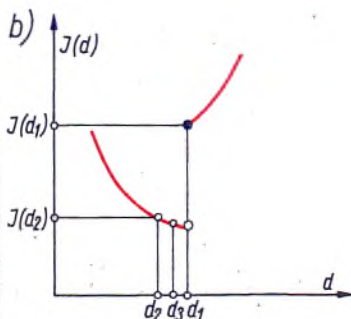
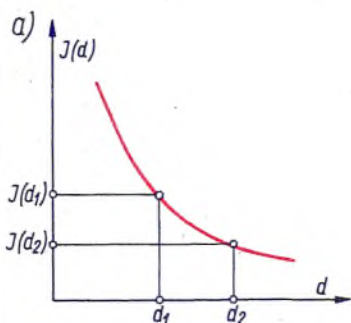
Może istnieć taka decyzja $\hat{d}$, że niewielkie odchyłki od niej powodują pogorszenie wskaźnika jakości, jakkolwiek w zbiorze decyzji dopuszczalnych istnieją inne decyzje zapewniające lepszą wartość wskaźnika jakości, niż $\hat{d}$. Mówimy wówczas, że $\hat{d}$ zapewnia optimum lokalne wskaźnika jakości (rys. 6.11B).

•6.12• Czy rozwiązanie zadania optymalizacji zawsze istnieje?

Nie zawsze. Rozwiązanie istnieje, jeżeli mamy do czynienia z problemem, w którym zbiorem decyzji dopuszczalnych jest skończony zbiór liczb całkowitych — wtedy liczba możliwych decyzji jest skończona, a zatem można spośród nich wybrać decyzję najlepszą.

W bardziej złożonych przypadkach rozwiązanie może nie istnieć. Zastanówmy się, co to znaczy, że rozwiązanie problemu nie istnieje. Jak wynika z definicji decyzji optymalnej, dowolnie mała odchyłka od rozwiązania powoduje pogorszenie wskaźnika jakości. W przykładach pokazanych na rys. 6.12A żaden z punktów nie może być rozwiązaniem, można bowiem wskazać punkt dowolnie mało różniący się od zadanego, a jednak zapewniający zmniejszenie wskaźnika jakości (a więc „lepsy”).

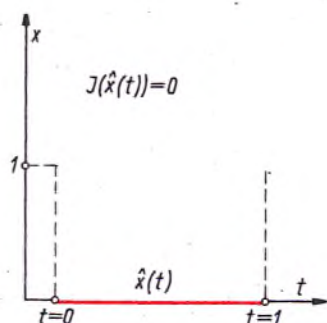
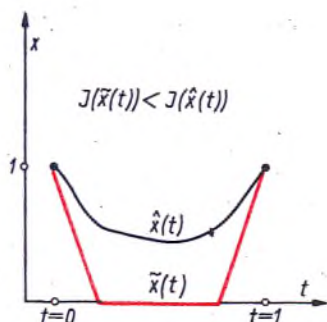
W przypadku zadań wariacyjnych (a więc gdy zbiorem de-



• 6.12A. Przykład zadań optymalizacji nie mających rozwiązania: a) dla każdej wartości d_1 , zawsze znajdzie się taka wartość d_2 (np. $d_1 + 1$), że wartość funkcji w punkcie d_2 będzie mniejsza niż w d_1 ; b) dla każdej decyzji d_2 , zawsze znajdzie się decyzję d_3 (np. $d_3 = \frac{1}{2}(d_1 + d_2)$), taka, że wartość funkcji w punkcie d_3 jest mniejsza niż w d_2

• 6.12B. Ilustracja rozwiązania zadania minimalizacji całki $\int_0^1 |x(t)| dt$ dla każdej decyzji $x(t)$ potrafimy wskazać decyzję $\hat{x}(t)$ zapewniającą mniejszą wartość całki (sytuacja analogiczna, jak na rys. 6.12A(a))

• 6.12C. Funkcja nieciągła zapewniająca minimum całki $\int_0^1 |x(t)| dt$



czyżby możliwych jest zbiór funkcji) sytuacja jest znacznie bardziej złożona. Istnienie lub nieistnienie rozwiązania może w istotny sposób zależeć od właściwości funkcji będącej rozwiązaniem.

Rozpatrzmy prosty przykład. Znaleźć funkcję $x(t)$ ciągłą, spełniającą warunki graniczne

$$x(0) = 1, \quad x(1) = 1$$

i zapewniającą minimum funkcjonału

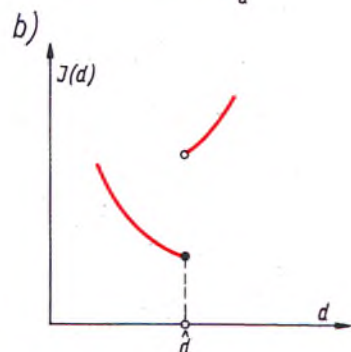
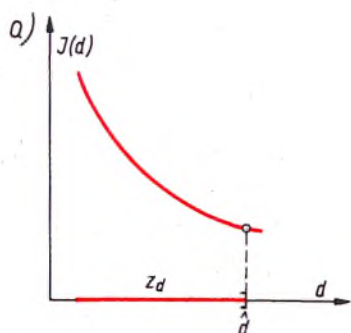
$$J = \int_0^1 |x(t)| dt$$

Tak sformułowane zadanie nie posiada rozwiązania. Przypuśćmy, że wybraliśmy pewną funkcję $\hat{x}(t)$ spełniającą warunki graniczne. Łatwo teraz znaleźć funkcję $\tilde{x}(t)$ zapewniającą mniejszą wartość funkcjonału, niż wartość przyjmowaną dla $\hat{x}(t)$ (rys. 6.12B). Wynika stąd nieistnienie rozwiązania. Jeżeli teraz zrezygnujemy z żądania, aby funkcja będąca rozwiązaniem była ciągłą, to rozwiązanie będzie istniało. Rozwiązaniem tym będzie funkcja nieciągła pokazana na rys. 6.12C.

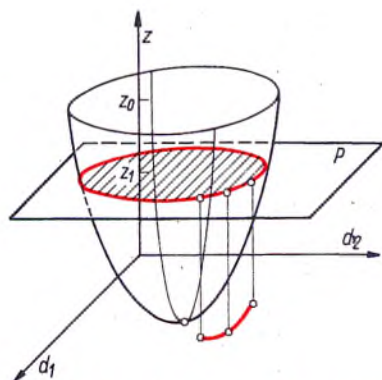
• 6.13. Czy można określić kiedy istnieje rozwiązanie zadania optymalizacji?

Można podać takie warunki, jednak w przypadku problemów wariacyjnych są one bardzo złożone. Podamy tu tylko warunek istnienia rozwiązania, gdy zbiorem decyzji możliwych jest zbiór liczb rzeczywistych. Warunek ten nosi nazwę **twierdzenia Weierstrassa**: Jeżeli zbiór decyzji dopuszczalnych jest domknięty i ograniczony, a wskaźnik jakości jest funkcją ciągłą, to istnieje rozwiązanie zadania optymalizacji. Zauważmy, że oba przykłady podane w pyt. 6.12 nie spełniały warunków twierdzenia: w pierwszym przykładzie zbiór decyzji dopuszczalnych był nieograniczony, a w drugim przypadku funkcja była nieciągła.

Należy zwrócić uwagę na sformułowanie twierdzenia — mówi ono, że o ile są spełnione pewne warunki, to istnieje rozwiązanie. Warunki te nie muszą być spełnione aby rozwiązanie istniało. Na rysunku 6.13 przedstawiono mody-



•6.13. Twierdzenie Weierstrassa jest warunkiem wystarczającym optymalności — funkcja może mieć minimum mimo niespełnienia warunków twierdzenia: a) nieograniczony zbiór decyzji dopuszczalnych; b) funkcja nieciągła



•6.14A. Sposób wyznaczania linii stałej wartości (poziomic) funkcji wielu zmiennych

fikację przykładów z pyt. 6.12 — w tym przypadku rozwiązanie obu zadań istnieje, mimo że warunki twierdzenia nie są spełnione. Tego typu warunki nazywamy warunkami wystarczającymi. Twierdzenie formułuje zatem warunki wystarczające istnienia ekstremum problemu optymalizacji z ograniczeniami.

•6.14. Jaka jest interpretacja geometryczna zadania programowania matematycznego?

W przypadku, gdy rozwiązując zadanie optymalizacji musimy określić wartość tylko jednej zmiennej decyzyjnej, interpretacja geometryczna problemu jest bardzo prosta i była stosowana poprzednio.

Zastanówmy się nad problemem, w którym występują dwie zmienne decyzyjne d_1 i d_2 .

Na płaszczyźnie (d_1, d_2) określimy zbiór decyzji dopuszczalnych (rys. 6.14A). Funkcja celu (wskaźnik jakości) jest funkcją dwóch zmiennych, nie można więc narysować jej na płaszczyźnie. Należy wprowadzić dodatkowo trzecią oś z , prostą do płaszczyzny (d_1, d_2) — jest to oś wartości funkcji. Każdemu punktowi o współrzędnych (d_1, d_2) jest przyporządkowany pewien punkt osi z , zatem wykresem funkcji jest w tym przypadku powierzchnia w przestrzeni trójwymiarowej. Ponieważ rysowanie wykresów przestrzennych jest kłopotliwe, wskaźnik jakości można określić za pomocą tzw. **linii poziomicowych** (poziomic). Wybierzmy jakiś punkt na osi z , (np. z_0) i poprowadźmy przez ten punkt płaszczyznę prostą do osi. Płaszczyzna ta przetnie powierzchnię będącą wykresem funkcji, przy czym punkty wspólne płaszczyzny i powierzchni będą tworzyły pewną krzywą. Przerzutując punkty tej krzywej na płaszczyznę (d_1, d_2) otrzymamy znów pewną krzywą. Krzywa ta ma następującą własność: jeżeli punkt (d_1, d_2) leży na tej krzywej, to wartość funkcji jest równa z_0 . Jest to zatem zbiór punktów stałej wartości funkcji. Zmieniając wartość z_0 otrzymamy inną krzywą; wykonując wiele takich rysunków otrzymamy całą rodzinę krzywych, przy czym z każdej z nich możemy przypisać wartość funkcji. Na takie wykresy nanosimy zbiór decyzji dopuszczalnych.

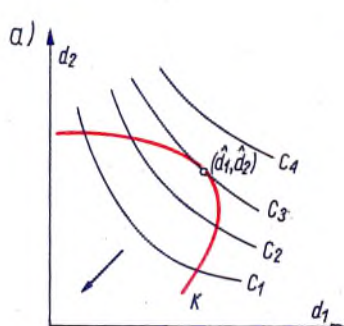
Można teraz łatwo wskazać rozwiązanie zadania optymalizacji. Na rysunku 6.14B przedstawiono 3 możliwe przypadki. Należy znaleźć punkt, który spełnia dwa warunki:

- leży w zbiorze decyzji dopuszczalnych,
- leży na linii poziomicowej odpowiadającej możliwie dużej wartości funkcji.

Punkt ten jest rozwiązaniem zadania optymalizacji.

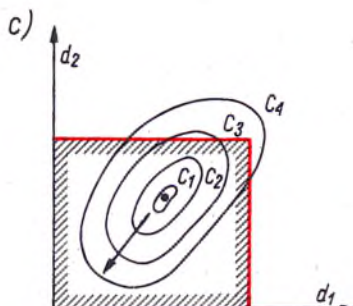
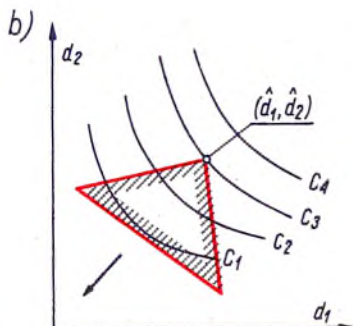
•6.15. Co to jest gradient funkcji i jaka jest jego interpretacja geometryczna?

Gradientem funkcji w danym punkcie nazywamy wektor, którego składowe są pochodnymi cząstkowymi funkcji względem poszczególnych zmiennych.



•6.14B• Interpretacja geometryczna rozwiązania zadania optymalizacji: a) zadanie z ograniczeniami równościowymi (krzywa K — zbiór decyzji dopuszczalnych); b) zadanie z ograniczeniami nierównościami — rozwiązanie na brzegu zbioru decyzji dopuszczalnych (ograniczenia aktywne); c) zadanie z ograniczeniami nierównościami — rozwiązanie wewnątrz zbioru decyzji dopuszczalnych:

C — linie stałej wartości funkcji; strzałką zaznaczono kierunek wzrostu funkcji



Rozpatrzmy właściwości takiego wektora w przypadku dwuwymiarowym. Ustalmy pewien punkt $(\hat{d}_1, \hat{d}_2)$. Jeżeli rozpatrujemy właściwości funkcji w małym otoczeniu tego punktu, to linie poziomicowe funkcji (linie stałej wartości) można traktować z dobrym przybliżeniem jako proste równoległe. Postawimy następujące pytanie: w jakim kierunku należy się przesunąć z punktu $(\hat{d}_1, \hat{d}_2)$ do punktu $(\hat{d}_1 + \delta, \hat{d}_2)$ odległego o δ tak, aby uzyskać największy przyrost funkcji. Odpowiedź łatwo znaleźć posługując się interpretacją geometryczną rys. 6.15. Należy poruszać się po prostej prostopadłej do poziomicy (a właściwie do prostej prostopadłej do stycznej do poziomicy) w punkcie $(\hat{d}_1, \hat{d}_2)$. Ponieważ rozpatrujemy małe otoczenie punktu $(\hat{d}_1, \hat{d}_2)$, to funkcję można w tym otoczeniu przybliżyć funkcją liniową o postaci

$$f = a_1 d_1 + a_2 d_2$$

Linie poziomicowe funkcji są więc opisane równaniem

$$a_1 d_1 + a_2 d_2 = C$$

gdzie C — pewna liczba.

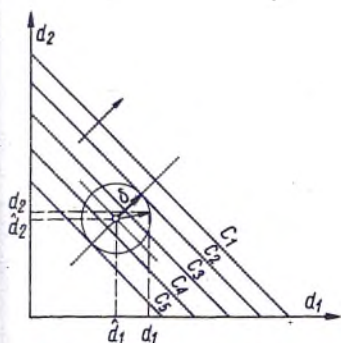
Łatwo zauważyć, że

$$a_1 = \frac{\partial f}{\partial d_1}, \quad a_2 = \frac{\partial f}{\partial d_2}$$

Tak więc równanie linii poziomicowej można zapisać

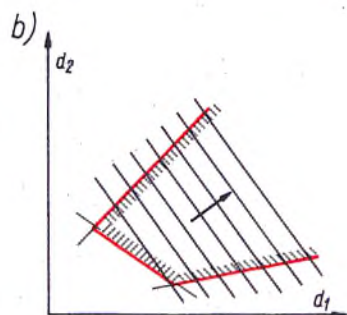
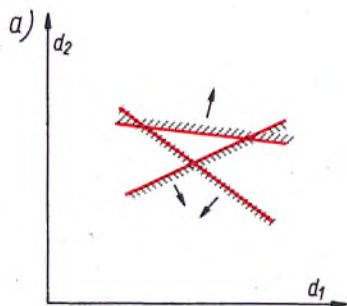
$$\frac{\partial f}{\partial d_1} d_1 + \frac{\partial f}{\partial d_2} d_2 = C$$

Wektor o składowych $\frac{\partial f}{\partial d_1}$, $\frac{\partial f}{\partial d_2}$ jest prostopadły do prostej określonej powyższym równaniem. A więc **gradient funkcji jest prostopadły do jej linii poziomicowej**. Można łatwo wykazać, że jego **zwrot wskazuje kierunek największego wzrostu funkcji**.

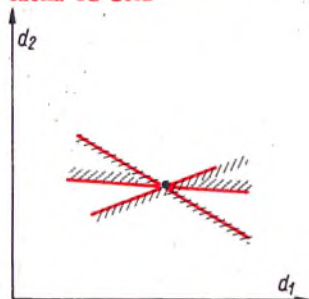


•6.15• Interpretacja geometryczna wektora gradientu

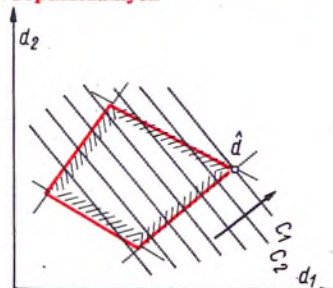
•6.16• Jakie są właściwości rozwiązań zadania programowania liniowego?



• 6.16A • Sytuacje, w których nie istnieje rozwiązanie zadania programowania liniowego: a) zbiór pusty decyzji dopuszczalnych; b) funkcja nieograniczona od dołu



• 6.16B • Jednopunktowy zbiór decyzji dopuszczalnych



• 6.16C • Zbiór decyzji dopuszczalnych jest ograniczonym wielobokiem

Wiadomo, że w najprostszym przypadku (2 wymiarowym) zadanie programowania liniowego ma postać: znaleźć minimum funkcji

$$f(x) = \alpha_1 x_1 + \alpha_2 x_2$$

przy ograniczeniach

$$\beta_{11} x_1 + \beta_{12} x_2 \leq d_1$$

$$\beta_{21} x_1 + \beta_{22} x_2 \leq d_2$$

$$\vdots \quad \quad \quad \vdots$$

$$\beta_{N1} x_1 + \beta_{N2} x_2 \leq d_N$$

(6.16.1)

Z rozważań przeprowadzonych poprzednio wynika, że zbiór decyzji dopuszczalnych jest wnętrzem pewnego wieloboku. Bokami jego są odcinki prostych o równaniach powstałych z zastąpienia znaku nierówności znakiem równości w równaniu (6.16.1).

Mogą mieć miejsce cztery przypadki:

— Zbiór decyzji dopuszczalnych jest pusty (rys. 6.16A(a)). Oznacza to, że nie istnieje ani jedna decyzja spełniająca wszystkie nierówności wchodzące w skład ograniczeń. Oczywiście w tym przypadku nie istnieje rozwiązanie zadania programowania liniowego.

— Zbiór decyzji składa się z jednego punktu, tzn. istnieje dokładnie jeden punkt spełniający wszystkie ograniczenia (rys. 6.16B). W tym przypadku trudno mówić o optymalizacji — rozwiązaniem jest po prostu ten jeden punkt, wynikający z równania więzów.

— Zbiór decyzji dopuszczalnych jest ograniczony. Wtedy zawsze istnieje rozwiązanie zadania optymalizacji. Ponieważ funkcja minimalizowana jest liniowa, to jej poziomicie (linie stałej wartości) są prostymi. Stąd rys. 6.16C wynika ważny wniosek: rozwiązanie zadania programowania liniowego (o ile istnieje) leży zawsze w wierzchołku lub na krawędzi wieloboku tworzącego zbiór decyzji dopuszczalnych.

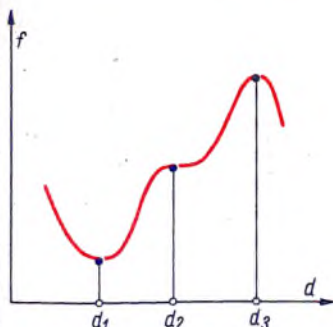
Zbiór decyzji dopuszczalnych jest nieograniczony (rys. 6.16A(b)). W tym przypadku rozwiązanie zadania programowania liniowego może nie istnieć, tzn. można wskazać decyzje dopuszczalne zapewniające dowolnie małe wartości funkcji celu.

• 6.17 • Co to są zadania optymalizacji bez ograniczeń i jak się je rozwiązuje?

Zadania optymalizacji nazywamy zadaniami bez ograniczeń w przypadku, gdy zbiorem decyzji dopuszczalnych a zarazem zbiorem decyzji możliwych jest zbiór liczb rzeczywistych.

Dla uproszczenia rozważmy przypadek, gdy można wybrać wartości dwóch zmiennych decyzyjnych $\hat{d}_1$ i $\hat{d}_2$.

Założmy, że w punkcie $(\hat{d}_1, \hat{d}_2)$ funkcja celu (wskaźnik jakości) $F(d_1, d_2)$ ma minimum lokalne. Oznacza to, że spełniona jest nierówność



• 6.17A • Zerowanie się pochodnej funkcji jest warunkiem koniecznym optymalności. W punktach d_1 i d_2 pochodna zeruje się, mimo to tylko w jednym z nich (d_1) istnieje minimum. W punkcie d_3 istnieje maksimum

$$F(\hat{d}_1, \hat{d}_2) \leq F(\hat{d}_1 + \varepsilon \cdot \Delta d_1, \hat{d}_2 + \varepsilon \cdot \Delta d_2) \quad (6.17.1)$$

We wzorze powyższym ε jest pewną małą liczbą, a Δd_1 , Δd_2 — odchyłkami od decyzji optymalnej. Korzystając ze wzoru Taylora możemy napisać

$$\begin{aligned} F(\hat{d}_1 + \varepsilon \cdot \Delta d_1, \hat{d}_2 + \varepsilon \cdot \Delta d_2) = & F(\hat{d}_1, \hat{d}_2) + \frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_1} \varepsilon \Delta d_1 + \\ & + \frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_2} \varepsilon \Delta d_2 + \frac{1}{2} \varepsilon^2 \frac{\partial^2 F(\hat{d}_1, \hat{d}_2)}{\partial d_1^2} \Delta^2 d_1 + \\ & + 2 \frac{\partial^2 F(\hat{d}_1, \hat{d}_2)}{\partial d_1 \partial d_2} \Delta d_1 \cdot \Delta d_2 + \frac{\partial^2 F(\hat{d}_1, \hat{d}_2)}{\partial d_2^2} \Delta^2 d_2 \end{aligned}$$

Z nierówności (6.17.1) wynika, że

$$\begin{aligned} \frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_1} \Delta d_1 + \frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_2} \Delta d_2 + \\ + \frac{1}{2} \varepsilon \left(\frac{\partial^2 F(\hat{d}_1, \hat{d}_2)}{\partial d_1^2} \Delta^2 d_1 + \dots \right) \geq 0 \end{aligned}$$

przy czym nierówność ta ma miejsce dla dowolnych wartości Δd_1 , Δd_2 i ε . Załóżmy, że Δd_1 , Δd_2 są ustalone. Wówczas nierówność jest spełniona dla dowolnie małego ε . Jeżeli ε jest dostatecznie małe można pominąć składniki zawierające drugie pochodne

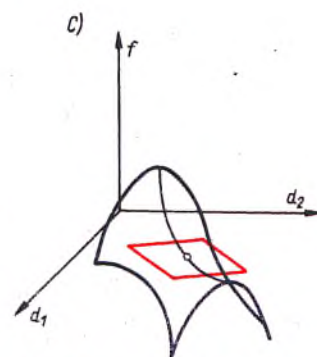
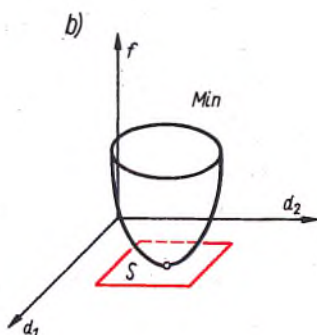
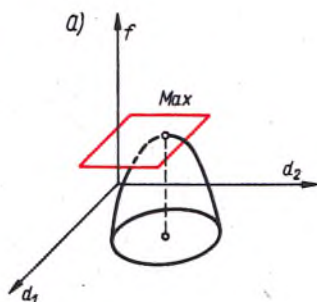
$$\frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_1} \Delta d_1 + \frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_2} \Delta d_2 \geq 0 \quad (6.17.2)$$

Nierówność ta jest spełniona dla dowolnych Δd_1 i Δd_2 tylko wtedy, gdy

$$\frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_1} = \frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_2} = 0$$

Tak więc, jeżeli $(\hat{d}_1, \hat{d}_2)$ jest rozwiązaniem zadania optymalizacji, to pochodne cząstkowe funkcji celu zerują się w tym punkcie. Przypomnimy, że w przypadku funkcji jednej zmiennej warunek ten oznacza, iż styczna do wykresu funkcji w punkcie optymalnym jest równoległa do osi x (rys. 6.17A). Podobnie interpretować można warunek (6.17.2) — oznacza on, że płaszczyzna styczna do powierzchni będącej wykresem funkcji jest równoległa do płaszczyzny (d_1, d_2) . Wiadomo, że spełnienie tego warunku nie wystarcza, aby w punkcie $(\hat{d}_1, \hat{d}_2)$ istniało ekstremum. Na rysunku 6.17B pokazano sytuację, gdy prosta lub płaszczyzna styczna są równoległe do osi lub płaszczyzny zmiennej niezależnej, a mimo to funkcja nie ma w tym punkcie ekstremum.

Stąd wynika ważny wniosek: chcąc znaleźć rozwiązanie zadania optymalizacji bez ograniczeń należy znaleźć wszystkie punkty, w których zerują się pochodne cząstkowe funkcji celu. Wśród tych punktów istnieją być może takie punkty, w których funkcja ma ekstremum.



•6.17B• Zerowanie się wektora gradientu jest warunkiem koniecznym optymalności

•6.18• Kiedy z zerowania się pochodnej wynika istnienie ekstremum w danym punkcie?

Dla uproszczenia rozpatrzmy znów przypadek funkcji dwóch zmiennych. Załóżmy, że w punkcie $(\hat{d}_1, \hat{d}_2)$ pochodne cząstkowe funkcji zerują się. Z zależności (6.17.2) z pytania poprzedniego wynika, że wówczas jest spełniony warunek

$$\frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_1} \Delta d_1 + \frac{\partial F(\hat{d}_1, \hat{d}_2)}{\partial d_2} \Delta d_2 + \frac{1}{2} \epsilon \left(\frac{\partial^2 F}{\partial d_1^2} \Delta^2 d_1 + 2 \frac{\partial^2 F}{\partial d_1 \partial d_2} \Delta d_1 \Delta d_2 + \frac{\partial^2 F}{\partial d_2^2} \Delta^2 d_2 \right) \geq 0$$

przy czym zachodzi on dla dowolnych $\Delta d_1, \Delta d_2$. Jeżeli oznaczymy

$$a_{11} = \frac{\partial^2 F}{\partial d_1^2}, \quad a_{12} = \frac{\partial^2 F}{\partial d_1 \partial d_2}, \quad a_{22} = \frac{\partial^2 F}{\partial d_2^2}$$

to warunki istnienia ekstremum można zapisać

$$a_{11} a_{22} - a_{12}^2 > 0 \quad \begin{cases} a_{11} > 0 & \text{— minimum} \\ a_{11} < 0 & \text{— maksimum} \end{cases}$$

$$a_{11} a_{22} - a_{12}^2 < 0 \quad \text{nie ma ekstremum.}$$

Przypadek, gdy któraś z nierówności jest spełniona równościowo wymaga specjalnego zbadania (trzeba badać zachowanie się wyższych pochodnych).

•6.19• W jaki sposób rozwiązuje się zadanie optymalizacji w przypadku, gdy zbiorem decyzji możliwych jest zbiór funkcji?

Rozważmy najpierw przypadek, gdy nie ma ograniczeń. Uwzględnienie ograniczeń jest w takich zadaniach na ogół dosyć skomplikowane. Sformułujemy następujący problem optymalizacji:

Znaleźć funkcję $x(t)$ dwukrotnie różniczkowalną, spełniającą warunki graniczne

$$x(a) = x_a, \quad x(b) = x_b$$

i zapewniającą minimum funkcjonału

$$J = \int_a^b F(x(t), \dot{x}(t), t) dt$$

Przypuśćmy, że znaleźliśmy funkcję $\hat{x}(t)$ będącą rozwiązaniem sformułowanego zadania. Porównamy wartości funkcjonału otrzymane dla funkcji w pewnym sensie „bliskich” $\hat{x}(t)$, z wartością otrzymaną dla $\hat{x}(t)$. Funkcje „bliskie” $\hat{x}(t)$ zapisujemy w postaci $x(t) = \hat{x}(t) + \epsilon \delta x(t)$

Funkcję $\delta x(t)$ nazywamy wariacją, powinna ona spełniać dodatkowo warunki: $\delta x(a) = 0, \delta x(b) = 0$. Warunki takie zapewniają, że funkcja $x(t)$ spełnia warunki graniczne.

Obliczymy wartość funkcjonału dla funkcji $x(t)$

$$J(\varepsilon) = \int_a^b F(x + \varepsilon \delta x, \dot{x} + \varepsilon \delta \dot{x}, t) dt$$

Jeżeli $\delta x(t)$ jest ustalona, to wartość funkcjonału zależy tylko od ε (jest to liczba), a zatem otrzymujemy funkcję jednej zmiennej. Odejmując wartości funkcjonału dla $x(t)$ i $\hat{x}(t)$ otrzymamy

$$\frac{J(\varepsilon) - J(0)}{\varepsilon} = \int_a^b \frac{F(x + \varepsilon \delta x, \dot{x} + \varepsilon \delta \dot{x}, t) - F(x, \dot{x}, t)}{\varepsilon} dt$$

Wyrażenie po lewej stronie znaku równości jest ilorazem różnicowym dla pochodnej $\frac{dJ}{d\varepsilon}$; wyrażenie pod całką można przepisać następująco:

$$\delta \dot{x} \frac{F(x + \varepsilon \delta x, \dot{x} + \varepsilon \delta \dot{x}, t) - F(x + \varepsilon \delta x, \dot{x}, t)}{\varepsilon \delta \dot{x}} + \\ + \frac{F(x + \varepsilon \delta x, \dot{x}, t) - F(x, \dot{x}, t)}{\varepsilon \delta x} \delta x$$

Pierwszy składnik sumy jest ilorazem różnicowym dla pochodnej $\frac{\partial F}{\partial \dot{x}}$, natomiast drugi składnik — dla pochodnej $\frac{\partial F}{\partial x}$.

$$\text{Stąd } \frac{dJ}{d\varepsilon} = \int_a^b \left(\frac{\partial F}{\partial \dot{x}} \delta \dot{x} + \frac{\partial F}{\partial x} \delta x \right) dt$$

Szukamy teraz takiej wartości ε , która zapewnia minimum funkcji $J(\varepsilon)$. Jak wiadomo, wartość tę określić można przyrównując do zera pochodną $\frac{dJ}{d\varepsilon}$. Wynika stąd, że jeżeli $\hat{x}(t)$ jest rozwiązaniem problemu optymalizacji, to spełniony jest warunek

$$\int_a^b \left(\frac{\partial F(\hat{x}, \hat{\dot{x}}, t)}{\partial \dot{x}} \delta \dot{x} + \frac{\partial F(\hat{x}, \hat{\dot{x}}, t)}{\partial x} \delta x \right) dt = 0 \quad (6.19.1)$$

Przekształcimy pierwszy składnik sumy stosując całkowanie przez części. Otrzymamy wówczas

$$\int_a^b \frac{\partial F}{\partial \dot{x}} \delta \dot{x} dt = \frac{\partial F(\hat{x}(t), \hat{\dot{x}}(t), t)}{\partial x} \delta x \Big|_a^b - \int_a^b \frac{d}{dt} \left(\frac{\partial F}{\partial \dot{x}} \right) \delta x dt$$

przy czym pierwsze dwa składniki po prawej stronie znaku równości

$$\frac{\partial F(\hat{x}(a), \hat{\dot{x}}(a), a)}{\partial \dot{x}} \delta x(a) + \frac{\partial F(\hat{x}(b), \hat{\dot{x}}(b), b)}{\partial \dot{x}} \delta x(b)$$

są równe zeru, gdyż $\delta x(a) = \delta x(b) = 0$

Wynika stąd, że równość (6.19.1) jest równoważna równości

$$\int_a^b \left(\frac{d}{dt} \left(\frac{\partial F}{\partial \dot{x}} \right) - \frac{\partial F}{\partial x} \right) \delta x(t) dt = 0$$

Ponieważ funkcję $\delta x(t)$ wybrano dowolnie, a równość powyższa powinna zachodzić dla dowolnych $\delta x(t)$, to jest ona równoważna warunkowi

$$\frac{d}{dt} \left(\frac{\partial F}{\partial \dot{x}} \right) - \frac{\partial F}{\partial x} = 0$$

Równanie to nazywamy równaniem **Eulera-Lagrange'a**. Jeżeli funkcja $\hat{x}(t)$ jest rozwiązaniem zadania optymalizacji, to spełnia równanie Eulera-Lagrange'a. Podkreślić należy, że z faktu iż funkcja spełnia równanie Eulera-Lagrange'a nie wynika, że jest ona rozwiązaniem zadania optymalizacji. Równanie to umożliwia tylko określenie funkcji podejrzanych o to, że są rozwiązaniami.

•6.20• Jak rozwiązuje się zadanie optymalizacji w przypadku istnienia ograniczeń równościowych (więzów)?

Założmy, że zbiorem decyzji możliwych analizowanego zadania optymalizacji jest zbiór liczb rzeczywistych. Zadanie polega na znalezieniu minimum funkcji

$$J(d_1, d_2)$$

przy ograniczeniach

$$g(d_1, d_2) = b$$

W prostych przypadkach można to ograniczenie rozwikłać, tzn. wyrazić d_1 przez d_2 (lub na odwrót) tzn. podać zależność

$$d_1 = h(d_2, b)$$

Otrzymamy wówczas funkcję jednej zmiennej d_2

$$J(h(d_2, b), d_2)$$

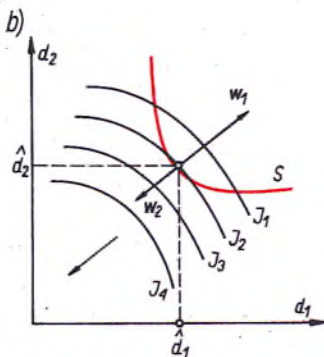
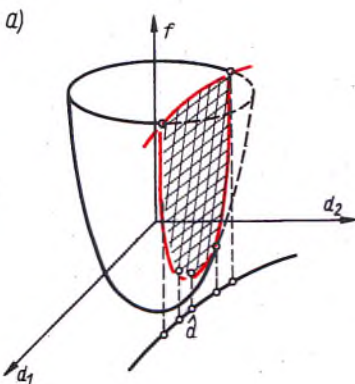
której minimum można znaleźć w prosty sposób.

Rozpatrzmy jeszcze raz interpretację graficzną zadania optymalizacji (rys. 6.20A). Poruszając się wzdłuż krzywej S określonej więzami obserwujemy wartości funkcji J . Rozwiązaniem jest punkt, który leży na krzywej i taki, w którym krzywa określona równaniem więzów jest styczna do linii stałej wartości odpowiadającej możliwie najmniejszej wartości funkcji. Krzywe są styczne, gdy kierunki wektorów normalnych w_1 i w_2 pokrywają się. Mamy zatem dwie krzywe o równaniach

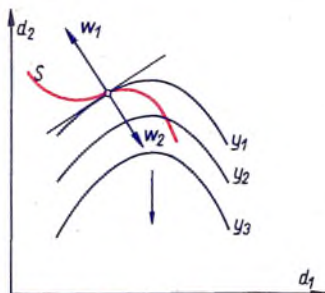
$$J(d_1, d_2) = c$$

$$g(d_1, d_2) = b$$

Warunek styczności można zapisać następująco:



•6.20A• Interpretacja geometryczna rozwiązania zadania optymalizacji z ograniczeniami równościowymi



• 6.20B • Styczność linii poziomowej funkcji J do krzywej S będącej zbiorem decyzji dopuszczalnych jest warunkiem koniecznym optymalności. Styczność nie gwarantuje, że w danym punkcie istnieje minimum funkcji

$$\frac{\partial J}{\partial d_1} = \alpha \frac{\partial g}{\partial d_1}, \quad \frac{\partial J}{\partial d_2} = \alpha \frac{\partial g}{\partial d_2}$$

gdzie α jest pewną liczbą.

Należy zatem szukać takiego punktu (d_1, d_2) , aby spełnione były powyższe nierówności.

Oczywiście równości te nie są warunkiem wystarczającym optymalności. Można podać przykłady, że mimo styczności krzywych w danym punkcie nie ma ekstremum (rys. 6.20B); równania te umożliwiają jedynie znalezienie wszystkich punktów podejrzanych o ekstremum.

Zauważmy, że jeżeli wprowadzimy funkcję

$$L(d_1, d_2, \alpha) = J(d_1, d_2) + \alpha(g(d_1, d_2) - b)$$

i poszukiwać będziemy punktu, w którym zerują się pochodne tej funkcji względem d_1, d_2, α , to otrzymamy

$$\frac{\partial L}{\partial d_1} = \frac{\partial J}{\partial d_1} + \alpha \frac{\partial g}{\partial d_1}$$

$$\frac{\partial L}{\partial d_2} = \frac{\partial J}{\partial d_2} + \alpha \frac{\partial g}{\partial d_2}$$

$$\frac{\partial L}{\partial \alpha} = g(d_1, d_2) - b$$

Wynika stąd, że zadanie poszukiwania ekstremum funkcji przy ograniczeniach równościowych można sprowadzić do zadania szukania punktu stacjonarnego pewnej funkcji rozszerzonej; zadanie to jest już zadaniem bez ograniczeń. Funkcję tę nazywamy **funkcją Lagrange'a** a parametr α — **mnożnikiem Lagrange'a**.

Podkreślić należy, że obie metody (tzn. rozwikłanie ograniczeń i metoda mnożników Lagrange'a) są w pełni równoważne i prowadzą do tego samego rozwiązania. Wybór metody zależy od specyfiki zadania i uznania rozwiązującego.

I tak np. przy więzach liniowych korzystne jest stosowanie rozwikływania ograniczeń ze względu na łatwość dokonania tej operacji. Istnieje duża grupa metod obliczeniowych (np. metoda Simpleks) stosujących tę technikę.

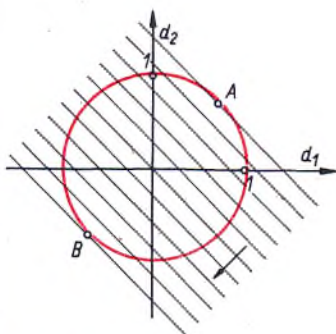
Z kolei przy obliczeniach „ręcznych”, gdy zależy nam na otrzymaniu rozwiązań analitycznych, korzystne jest stosowanie metody mnożników Lagrange'a (nawet w przypadku liniowych więzów). Stosowanie tej metody znacznie zmniejsza pracochłonność rachunków.

Przykład zadania minimalizacji z ograniczeniami równościowymi.

Zadanie: Znaleźć minimum funkcji $d_1 + d_2$, przy ograniczeniach

$$d_1^2 + d_2^2 = 1.$$

Interpretację graficzną zadania pokazuje rys. 6.20C.



• 6.20C •

Wprowadzamy funkcję Lagrange'a

$$L(d_1, d_2, \alpha) = d_1 + d_2 + \alpha(d_1^2 + d_2^2 - 1)$$

i poszukujemy punktu stacjonarnego:

$$\left. \begin{aligned} \frac{\partial L}{\partial d_1} &= 1 + 2\alpha d_1 = 0 \\ \frac{\partial L}{\partial d_2} &= 1 + 2\alpha d_2 = 0 \\ -\frac{\partial L}{\partial \alpha} &= d_1^2 + d_2^2 - 1 = 0 \end{aligned} \right\}$$

Z dwóch pierwszych równań wynika

$$\alpha = -\frac{1}{2d_1}, \quad \alpha = -\frac{1}{2d_2}$$

zatem $d_1 = d_2 = d$.

Podstawiając do trzeciego równania otrzymamy

$$d^2 + d^2 = 1, \text{ czyli } 2d^2 = 1$$

$$\text{stąd } d = \pm \frac{1}{\sqrt{2}}$$

Otrzymujemy zatem dwa punkty „podejrzane” o minimum: $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$. Obliczymy wartość funkcji w tych punktach — w pierwszym przypadku jest ona równa $\frac{2}{\sqrt{2}}$, w drugim — $-\frac{2}{\sqrt{2}}$. Wynika stąd, że funkcja posiada minimum w punkcie $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$.

·6.21· W jaki sposób rozwiązuje się zadanie optymalizacji z ograniczeniami znaku zmiennych decyzyjnych?

Ograniczenia tego typu bardzo często występują w praktycznych zadaniach optymalizacji. Rozpatrzmy zadanie: znaleźć minimum funkcji $F(d)$ przy ograniczeniu $d \geq 0$.

Trzy możliwe przypadki pokazano na rys. 6.21A(a, b, c). Wynika stąd następujący sposób rozwiązywania zadania:

— Obliczyć pochodną $\frac{\partial F}{\partial d}$, znaleźć $\hat{d}$, dla której pochodna zeruje się;

— Jeżeli $\hat{d} > 0$, to jest rozwiązaniem zadania;

— Jeżeli $\hat{d} \leq 0$, to rozwiązaniem zadania jest $\hat{d} = 0$.

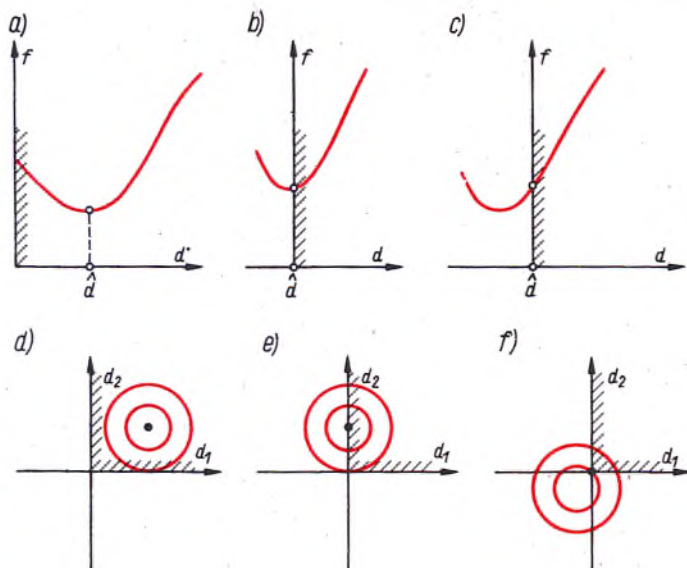
Zajmiemy się teraz funkcją dwóch zmiennych.

Możliwe są następujące przypadki:

— Minimum znajduje się wewnątrz obszaru dopuszczalnego;

— Minimum znajduje się na jednym z ograniczeń $d_1 = 0$ lub $d_2 = 0$ (jeżeli ograniczenie jest spełnione równościowo to nazywane jest aktywnym).

•6.21A• Interpretacja geometryczna rozwiązania zadania minimalizacji funkcji przy ograniczeniach znaku zmiennych: a, b, c) — minimalizacja funkcji jednej zmiennej; d, e, f) — minimalizacja funkcji dwóch zmiennych



— Minimum znajduje się na obu ograniczeniach $d_1 = 0$, $d_2 = 0$. Z rysunków 6.21A(d, e, f) oraz z rozważań poprzednich wynika następujący wniosek:

Jeżeli w punkcie będącym rozwiązaniem zadania któregoś z ograniczeń jest aktywne (spełnione równościowo), to pochodna cząstkowa względem tej zmiennej jest w tym punkcie dodatnia. Analitycznie warunki te można zapisać

$$\left. \begin{aligned} \frac{\partial F(\hat{d})}{\partial d_i} &\geq 0 \\ \frac{\partial F(\hat{d})}{\partial d_i} \hat{d}_i &= 0 \\ \hat{d}_i &\geq 0 \end{aligned} \right\} \quad (\text{dla wszystkich zmiennych})$$

Tak więc, jeżeli punkt $\hat{d}$ jest rozwiązaniem zadania to zapewnia spełnienie tych warunków.

Przykład rozwiązania zadania z ograniczeniami znaku zmiennych.

Zadanie: (rys. 6.21B):

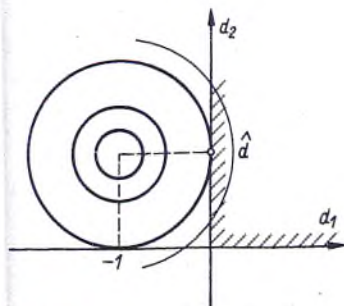
Znaleźć minimum funkcji $(d_1+1)^2 + (d_2-1)^2$ przy ograniczeniach $d_1 \geq 0$, $d_2 \geq 0$.

Obliczamy pochodne funkcji F :

$$\frac{\partial F}{\partial d_1} = 2(d_1+1)$$

$$\frac{\partial F}{\partial d_2} = 2(d_2-1)$$

Warunki optymalności można zapisać następująco:



•6.21B•

d_1	> 0	$= 0$	> 0	$= 0$
d_2	> 0	> 0	$= 0$	$= 0$
Sytuacja	1	2	3	4

$$d_1 + 1 \geq 0 \quad (6.21.1)$$

$$d_2 - 1 \geq 0 \quad (6.21.2)$$

$$d_1(d_1 + 1) = 0 \quad (6.21.3)$$

$$d_2(d_2 - 1) = 0 \quad (6.21.4)$$

$$d_1 \geq 0, \quad d_2 \geq 0 \quad (6.21.5)$$

Przedyskutujemy te warunki. Możliwe sytuacje zapisujemy w tabl. 6.21.

Należy sprawdzić, dla której sytuacji spełnione są warunki (6.21.1) ... (6.21.5).

Sytuacja 1:

d_1 i d_2 są większe od zera, zatem z równań (6.21.3), (6.21.4) wynika $d_1 = -1$, $d_2 = 1$ — nie są spełnione nierówności z pierwszej kolumny tabeli.

Sytuacja 2:

$$d_1 = 0, \quad d_2 > 0$$

Sprawdzamy kolejno:

(6.21.1) — nierówność spełniona, bo $0 + 1 \geq 0$;

$d_2 > 0$, a zatem z (6.21.4) wynika, że $d_2 = 1$

(6.21.2) — nierówność spełniona, bo $1 - 1 \geq 0$

(6.21.3) — równość spełniona.

(6.21.4) — równość spełniona.

Wynika stąd, że warunki konieczne optymalności są spełnione dla $d_1 = 0$, $d_2 = 1$.

Łatwo sprawdzić, że sytuacje 3 i 4 nie zapewniają jednoczesnego spełnienia wszystkich warunków (6.21.1) ... (6.21.5).

•6.22• W jaki sposób rozwiązuje się zadanie optymalizacji z ograniczeniami nierównościowymi?

Rozpatrzmy następujące zadanie:

Znaleźć minimum funkcji $F(d_1, d_2)$ przy ograniczeniach

$$g(d_1, d_2) - b \leq 0, \quad h(d_1, d_2) - c \leq 0$$

Obszar dopuszczalny jest ograniczony krzywymi o równaniach

$$g(d_1, d_2) = b \text{ oraz } h(d_1, d_2) = c.$$

Podobnie jak w poprzednim przypadku możliwe są trzy sytuacje:

— Funkcja posiada minimum we wnętrzu obszaru dopuszczalnego;

— Jedno z ograniczeń jest spełnione równościowo;

— Oba ograniczenia są spełnione równościowo.

Zapiśzemy warunki analityczne, które musi spełniać funkcja minimalizowana.

W sytuacji pierwszej, gdy rozwiązanie leży wewnątrz obszaru dopuszczalnego, warunki konieczne optymalności sprowadzają się do żądania zerowania się obu pochodnych cząstkowych

$$\frac{\partial F}{\partial d_1} = \frac{\partial F}{\partial d_2} = 0$$

W sytuacji drugiej, gdy jedno z ograniczeń jest spełnione w sposób równościowy, mamy do czynienia z zadaniem minimalizacji funkcji przy ograniczeniu, np.

$$g(d_1, d_2) = b$$

Można więc zastosować metodę mnożników Lagrange'a, warunki optymalności przyjmą następującą postać:

$$\frac{\partial L}{\partial d_1} = \frac{\partial F}{\partial d_1} + \alpha \frac{\partial g}{\partial d_1} = 0$$

$$\frac{\partial L}{\partial d_2} = \frac{\partial F}{\partial d_2} + \alpha \frac{\partial g}{\partial d_2} = 0 \quad (6.22.1)$$

$$\frac{\partial L}{\partial \alpha} = g(d_1, d_2) - b = 0$$

Funkcja Lagrange'a ma postać: $L = F(d_1, d_2) + \alpha(g(d_1, d_2) - b)$

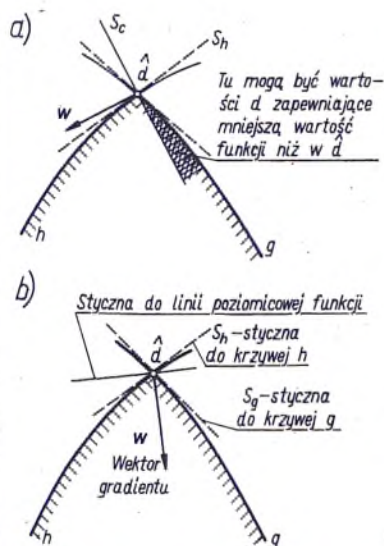
Dodatkowo należy wprowadzić ograniczenia znaku mnożnika Lagrange'a α .

Jeżeli rozwiązanie leży na brzegu obszaru dopuszczalnego, to poruszanie się w głąb tego obszaru powoduje wzrost funkcji. Wynika stąd, że wektor gradientu funkcji powinien być skierowany do wnętrza obszaru dopuszczalnego. Z kolei obszar dopuszczalny jest określony nierównością $g(d_1, d_2) - b \leq 0$, a więc wektor normalny stycznej do brzegu obszaru jest skierowany na zewnątrz obszaru dopuszczalnego. Wektor gradientu funkcji celu i wektor normalny do ograniczeń mają ten sam kierunek, ale przeciwny zwrot, zatem zestaw równości (6.22.1) należy uzupełnić nierównością $\alpha > 0$. Wynika to z faktu, że α jest współczynnikiem proporcjonalności między wektorem gradientu, a wektorem normalnym do ograniczeń, a wektory te mają przeciwny zwrot. Fakt ten zapisujemy

$$\frac{\partial F}{\partial d_1} = -\alpha \frac{\partial g}{\partial d_1}, \quad \frac{\partial F}{\partial d_2} = -\alpha \frac{\partial g}{\partial d_2}, \quad \alpha > 0$$

Nieco bardziej skomplikowany jest przypadek, gdy oba ograniczenia są aktywne w punkcie będącym rozwiązaniem. Z rysunku 6.22A widać, że przypadek taki może mieć miejsce tylko wtedy, gdy styczna do powierzchni stałej wartości jest zawarta między stycznymi do ograniczeń, a ponadto wektor gradientu funkcji jest skierowany do wnętrza obszaru dopuszczalnego. Warunki te, w postaci analitycznej są następujące:

$$-\frac{\partial F}{\partial d_1} = \alpha \frac{\partial g}{\partial d_1} + \beta \frac{\partial h}{\partial d_1}$$



6.22A. Interpretacja geometryczna warunków koniecznych optymalności dla zadania optymalizacji z ograniczeniami nierównościowymi: a) nie spełnione warunki konieczne optymalności; b) spełnione warunki konieczne w punkcie d

$\frac{\partial F}{\partial d_1} = \frac{\partial F}{\partial d_2} = 0$	$\frac{\partial F}{\partial d_1} + \alpha \frac{\partial g}{\partial d_1} = 0$	$\frac{\partial E}{\partial d_1} + \beta \frac{\partial h}{\partial d_1} = 0$	$\frac{\partial F}{\partial d_1} + \alpha \frac{\partial g}{\partial d_1} + \beta \frac{\partial h}{\partial d_1} = 0$
$g(d_1, d_2) - b < 0$	$\frac{\partial F}{\partial d_2} + \alpha \frac{\partial g}{\partial d_2} = 0$	$\frac{\partial F}{\partial d_2} + \beta \frac{\partial h}{\partial d_2} = 0$	$\frac{\partial F}{\partial d_2} + \alpha \frac{\partial g}{\partial d_2} + \beta \frac{\partial h}{\partial d_2} = 0$
$h(d_1, d_2) - c < 0$	$g(d_1, d_2) - b = 0$	$g(d_1, d_2) - b < 0$	$g(d_1, d_2) - b = 0$
	$h(d_1, d_2) - c < 0$	$h(d_1, d_2) - c = 0$	$h(d_1, d_2) - c = 0$
	$\alpha \geq 0$	$\beta \geq 0$	$\alpha \geq 0 \quad \beta \geq 0$

$$-\frac{\partial F}{\partial d_2} = \alpha \frac{\partial g}{\partial d_2} + \beta \frac{\partial h}{\partial d_2}$$

przy czym współczynniki α i β muszą być dodatnie. Podsumujemy teraz otrzymane rezultaty:

Jeżeli punkt o współrzędnych $(\hat{d}_1, \hat{d}_2)$ jest rozwiązaniem zadania optymalizacji, to mogą zachodzić przypadki przedstawione w tabl. 6.22. Zauważmy, że przypadki te zapisać można wprowadzając funkcję Lagrange'a o postaci

$$L(d_1, d_2, \alpha, \beta) = F(d_1, d_2) + \alpha(g(d_1, d_2) - b) + \beta(h(d_1, d_2) - c)$$

Wówczas warunki można zapisać następująco:

$$\frac{\partial L}{\partial d_1} = 0, \frac{\partial L}{\partial d_2} = 0, \quad g(d_1, d_2) - b \leq 0, \quad h(d_1, d_2) - c \leq 0$$

$$\alpha(g(d_1, d_2) - b) = 0, \quad \beta(h(d_1, d_2) - c) = 0$$

$$\alpha \geq 0, \quad \beta \geq 0$$

Liczby α i β nazywamy mnożnikami Lagrange'a, a zespół nierówności **warunkami Kuhna-Tuckera**.

Przykład rozwiązania zadania optymalizacji z ograniczeniami nierównościami.

Zadanie (rys. 6.22B):

Znaleźć minimum funkcji $F(d_1, d_2) = d_1 + d_2$, przy ograniczeniach

$$d_1^2 + d_2^2 \leq 1$$

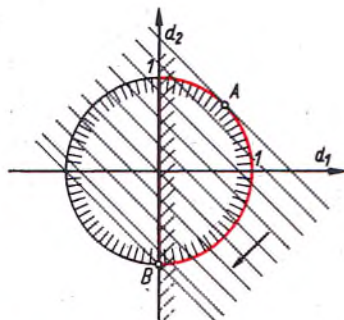
$$d_1 \geq 0 \quad (\text{czyli } -d_1 \leq 0)$$

Funkcja Lagrange'a dla tego zadania ma postać

$$L(d_1, d_2, \alpha, \beta) = d_1 + d_2 + \alpha(d_1^2 + d_2^2 - 1) - \beta d_1$$

Warunki Kuhna-Tuckera są następujące:

$$\frac{\partial L}{\partial d_1} = 1 + 2\alpha d_1 - \beta = 0$$



$$\frac{\partial L}{\partial d_2} = 1 + 2\alpha d_2 = 0$$

$$-d_1 \leq 0$$

$$d_1^2 - d_2^2 - 1 \leq 0$$

$$(d_1^2 + d_2^2 - 1)\alpha = 0$$

$$d_1 \beta = 0$$

$$\alpha \geq 0, \quad \beta \geq 0$$

Chcemy znaleźć punkty spełniające te warunki — będą to punkty „podejrzane” o ekstremum.

Rozważmy po kolei możliwe przypadki.

$$a) \alpha > 0, \quad \beta > 0$$

Z nierówności tych wynika, że

$$d_1 = 0, \quad d_1^2 + d_2^2 - 1 = 0, \quad \text{czyli } d_2 = \pm 1$$

Zbadamy, czy w punkcie (0, 1) spełnione są pozostałe warunki:

$$1 + 2\alpha \cdot 0 - \beta = 0, \quad \text{stąd } \beta = 1$$

$$1 + 2\alpha \cdot 1 = 0, \quad \text{stąd } \alpha = -\frac{1}{2} \text{ — niedopuszczalny znak } \alpha$$

Wynika stąd, że punkt (0, 1) nie może być podejrzany o ekstremum. Zbadamy punkt (0, -1)

$$1 + 2\alpha \cdot 0 - \beta = 0 \quad \text{stąd } \beta = 1$$

$$1 - 2\alpha = 0 \quad \text{stąd } \alpha = \frac{1}{2}$$

$$d_1^2 + d_2^2 - 1 = 0 + 1 - 1 = 0 \leq 0$$

$$(d_1 + d_2 - 1)\alpha = (0 + 1 - 1) \frac{1}{2} = 0$$

$$d_1 \beta = 0 \cdot 1 = 0$$

W punkcie (0, -1) może być ekstremum.

$$b) \alpha = 0, \quad \beta > 0$$

Łatwo sprawdzić, że ten i pozostałe warunki nie zapewniają spełnienia wszystkich warunków Kuhna-Tuckera.

· 6.23 · Czy można zastosować metodę mnożników Lagrange'a do rozwiązywania zadań wariacyjnych?

Można, chociaż w wielu przypadkach jest to dosyć trudne. Sposób stosowania tej metody zilustrujemy na przykładzie, odsyłając Czytelnika do podręczników rachunku wariacyjnego. Założmy, że mamy do rozwiązania następujący problem wariacyjny:

Znaleźć minimum funkcjonału

$$J = \int_0^T (x^2(t) + u^2(t)) dt$$

przy spełnionych równaniach różniczkowych więzów

$$\dot{x}(t) = u(t), \quad x(0) = 1$$

Funkcje $x(t)$, $u(t)$ — są w tym przypadku zmiennymi decyzyjnymi — należy wybrać je tak, aby uzyskać minimum funkcjonału (oraz zapewnić spełnienie ograniczeń).

Wprowadzamy funkcjonał Lagrange'a o następującej postaci

$$L(u, x, \alpha) = \int_0^T [x^2(t) + u^2(t) - \alpha(t) (\dot{x}(t) - u(t))] dt$$

Zauważmy, że mnożnik Lagrange'a $\alpha(t)$ jest teraz funkcją. Sposób postępowania jest identyczny, jak w przypadku zadania bez ograniczeń.

Założmy, że znaleźliśmy funkcje $\hat{x}(t)$, $\hat{u}(t)$ zapewniające minimum funkcjonału. Zmienimy trochę te funkcje, wprowadzając wariacje $\delta x(t)$, $\delta u(t)$. Wariacje te powinny spełniać warunek $\delta x(0) = 0$. Zmienione funkcje (tzw. „odchylone”) określimy następująco:

$$x(t) = \hat{x}(t) + \varepsilon \delta x(t), \quad u(t) = \hat{u}(t) + \varepsilon \delta u(t)$$

Podstawiając te „odchylone” funkcje do funkcjonału Lagrange'a uzależnimy wartość funkcjonału od parametru ε :

$$L(\varepsilon) = \int_0^T [(\hat{x} + \varepsilon \delta x)^2 + (\hat{u} + \varepsilon \delta u)^2 - \alpha(\hat{x} + \varepsilon \delta \dot{x} - \hat{u} - \varepsilon \delta u)] dt$$

Zauważmy, że para funkcji $\hat{x}(t)$, $\hat{u}(t)$ jest rozwiązaniem zadania, a więc spełnia ograniczenia

$$\dot{\hat{x}}(t) = \hat{u}(t)$$

zatem funkcję $L(\varepsilon)$ można przepisać następująco:

$$L(\varepsilon) = \int_0^T \{(\hat{x} + \varepsilon \delta x)^2 + (\hat{u} + \varepsilon \delta u)^2 - \alpha(\varepsilon \delta \dot{x} - \varepsilon \delta u)\} dt$$

Zbadajmy nieco dokładniej funkcję $L(\varepsilon)$.

Rozpisując odpowiednie wyrażenia otrzymamy

$$L(\varepsilon) = \varepsilon^2 \int_0^T \{(\delta x)^2 + (\delta u)^2\} dt + \varepsilon \int_0^T \{\hat{x} \delta x + \hat{u} \delta u - \alpha(\delta \dot{x} - \delta u)\} dt + \int_0^T \{\hat{x}^2 + \hat{u}^2\} dt$$

Ponieważ funkcje $\delta x(t)$, $\delta u(t)$, $\alpha(t)$ są ustalone, to $L(\varepsilon)$ jest funkcją kwadratową parametru ε . Jeżeli $(\hat{x}(t), \hat{u}(t))$ jest rozwiązaniem optymalnym, to dowolna zmiana tych funkcji zwiększa wartość funkcjonału jakości. Fakt ten ma miejsce wówczas, gdy spełniona jest równość

$$\int_0^T \{2\hat{x} \delta x + 2\hat{u} \delta u - \alpha(\delta \dot{x} - \delta u)\} dt = 0$$

W wyrażeniu tym usuniemy $\delta \dot{x}(t)$ całkując przez części

$$\int_0^T \alpha(t) \delta \dot{x}(t) dt = \alpha(T) \delta x(T) - \alpha(0) \delta x(0) - \int_0^T \dot{\alpha}(t) \delta x(t) dt$$

Zauważmy, że $\alpha(0) \delta x(0) = 0$, bo funkcja $\delta x(t)$ spełnia warunek $\delta x(0) = 0$ (lewy koniec trajektorii ustalony).

Otrzymujemy zatem

$$\int_0^T \{ (2\hat{x}(t) + \dot{\alpha}(t)) \delta x(t) + (2\hat{u}(t) + \alpha(t)) \delta u(t) \} dt + \alpha(T) \delta x(T) = 0$$

Jeżeli spełnione są warunki

$\alpha(T) = 0$, $\dot{\alpha}(t) = -2\hat{x}(t)$, $\hat{u}(t) = -\frac{1}{2}\alpha(t)$, to równość ta zachodzi dla dowolnych $\delta x(t)$ i $\delta u(t)$. Wynika stąd, że warunki powyższe są jednocześnie warunkami koniecznymi optymalności pary funkcji $(\hat{x}(t), \hat{u}(t))$.

Reasumując, jeżeli para funkcji $(\hat{x}(t), \hat{u}(t))$ jest rozwiązaniem optymalnym, to spełnia następujący zespół warunków:

1. $\dot{\hat{x}}(t) = \hat{u}(t)$
2. $\dot{\alpha}(t) = 2\hat{x}(t)$
3. $\hat{u}(t) = \frac{1}{2}\alpha(t)$

$$\alpha(T) = 0, \quad x(0) = 1$$

Wygodnie jest wprowadzić funkcję zwaną **hamiltonianem**. Ma ona postać

$$H(u, x, \alpha) = -(x^2 + u^2) + \alpha u$$

Za pomocą tej funkcji warunki optymalności zapisuje się szczególnie prosto

1. $\dot{\hat{x}} = \frac{\partial H(\hat{u}, \hat{x}, \alpha)}{\partial \alpha}$
2. $\dot{\alpha} = -\frac{\partial H(\hat{u}, \hat{x}, \alpha)}{\partial x}$

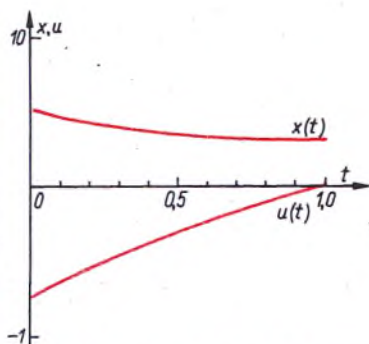
3. — zależność między $\hat{u}$, α , $\hat{x}$ ustala się na podstawie żądania, aby $\hat{u}$ zapewniało maksimum funkcji H w każdej chwili czasu. Istotnie, różniczkując tę funkcję względem u otrzymamy

$$\frac{\partial H}{\partial u} = -2u + \alpha$$

Ponieważ u ma zapewniać (zgodnie z zasadą maksimum) maksimum hamiltonianu, to powinien być spełniony warunek

$$\frac{\partial H}{\partial u} = 0, \quad \text{czyli } \hat{u} = \frac{1}{2}\alpha$$

Łatwo sprawdzić, że ostateczne rozwiązanie zadania ma postać (dla $T = 1$, $x(0) = \frac{1}{2}$), (rys. 6.23)



•6.23• Przebieg optymalnego sterowania i odpowiadająca mu trajektoria stanu

$$\dot{x}(t) = \frac{e^{1-t} + e^{t-1}}{2(e^{-1} + e)}, \quad \dot{u}(t) = \frac{-e^{1-t} + e^{t-1}}{e^{-1} + e}$$

·6.24· Czy posługując się metodą mnożników Lagrange'a można rozwiązać zadania wariacyjne z ograniczeniami?

Rozwiązanie takich zadań jest znacznie bardziej skomplikowane niż w poprzednim przypadku. Jedną z najlepszych metod rozwiązywania takich zagadnień jest tzw. **zasada maksimum** sformułowana przez Pontriagina. Podamy najprostszą wersję tej zasady.

Rozważmy następujący problem optymalizacji:

Znaleźć minimum funkcjonału

$$J = \int_0^T f_0(x(t), u(t), t) dt$$

przy ograniczeniach równościowych (wiązach)

$$\dot{x} = f(x, u, t), \quad x(0) = x_0$$

i ograniczeniach nierównościowych

$$u_{\min} \leq u(t) \leq u_{\max} \quad (6.24.1)$$

przy czym ostatnia nierówność ma być spełniona dla wszystkich chwil czasu z przedziału $[0, T]$.

Rozwiązanie

Wprowadzamy hamiltonian

$$H(x, u, \Psi, t) = -f_0(x, u, t) + \Psi f(x, u, t)$$

wówczas warunki konieczne optymalności są następujące:

$$\dot{x} = \frac{\partial H}{\partial \Psi} \quad x(0) = x_0$$

$$\dot{\Psi} = -\frac{\partial H}{\partial x} \quad \Psi(T) = 0$$

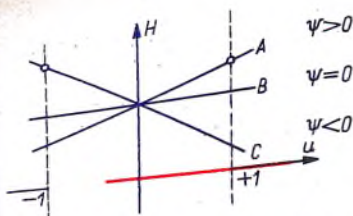
Zależność między x, u, Ψ ustala się na podstawie żądania, aby u zapewniało w każdej chwili czasu maksimum hamiltonianu na zbiorze określonym nierównościami (6.24.1).

Zauważmy podobieństwo tych warunków do warunków otrzymanych w zadaniu poprzednim — zasada maksimum różni się tylko tym, że szukanie maksimum hamiltonianu może być wykonywane na pewnym zbiorze określonym przez ograniczenie zmiennych decyzyjnych. Wprowadzenie innych ograniczeń, np.

$$x_{\min} \leq x(t) \leq x_{\max}$$

powoduje bardzo znaczną komplikację problemu; zadania takie nie będą tu rozważane.

·6.25· Jaki można podać przykład rozwiązywania zadania wariacyjnego z ograniczeniami nierównościowymi za pomocą zasady maksimum?



$$\hat{u} = \begin{cases} -1 & \text{gdy } \psi < 0 \\ \text{dowolne} & \text{gdy } \psi = 0 \\ +1 & \text{gdy } \psi > 0 \end{cases} \in [-1, 1]$$

• 6.25A • Rozwiązanie graficzne zadania maksymalizacji Hamiltonianu

Zadanie:

Znaleźć minimum funkcjonau

$$J = \int_0^T x^2(t) dt$$

przy spełnionych ograniczeniach równościowych (więzach)

$$\dot{x} = u, \quad x(0) = 1$$

i ograniczeniach nierównościowych

$$-1 \leq u(t) \leq 1$$

(6.25.1)

Wprowadzamy hamiltonian

$$H = -x^2 + \Psi u$$

i szukamy takiej wartości u , która zapewnia w każdej chwili maksimum hamiltonianu, przy czym należy do zbioru określonego nierównościami (6.25.1).

Mamy tu do czynienia z poszukiwaniem ekstremum funkcji jednej zmiennej przy ograniczeniach nierównościowych. Zadanie rozwiążemy graficznie (rys. 6.25A). Z rysunku wynika, że możliwe są trzy przypadki:

1. $\Psi < 0$

Wówczas maksimum H uzyskamy, jeżeli przyjmiemy $u = -1$

2. $\Psi > 0$

W tym przypadku maksimum H otrzymamy dla $u = +1$

3. $\Psi = 0$

W tym przypadku u może przyjmować dowolne wartości z przedziału $[-1, 1]$.

Wypiszemy pozostałe warunki optymalności:

$$\dot{x} = u, \quad x(0) = 1 \quad (6.25.2)$$

$$\dot{\Psi} = 2x, \quad \Psi(T) = 0 \quad (6.25.3)$$

Należy znaleźć funkcje $x(t)$, $\Psi(t)$ spełniające powyższe równania. Przypuśćmy, że w całym przedziale $[0, T]$ funkcja $\Psi(T)$ jest dodatnia. Zbadamy konsekwencje takiego przypuszczenia.

Z powyższych rozważań wynika, że sterowanie przyjmuje wartość $+1$. Pierwsze z równań ma zatem postać

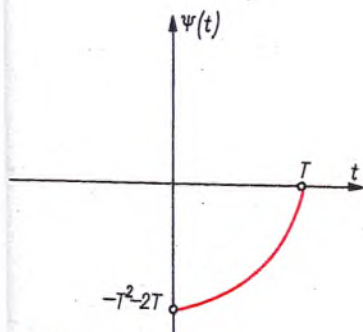
$$\dot{x} = 1$$

a rozwiązaniem jest funkcja

$$x(t) = t + 1$$

Wynika stąd, że funkcja Ψ jest następująca

$$\Psi(t) = 2 \left[\frac{t^2}{2} + t - \frac{T^2}{2} - T \right]$$



6.25B •

Funkcję tę przedstawiono na rys. 6.25B. Widać, że funkcja jest ujemna w przedziale $[0, T]$, a więc przypuszczenie nasze było niesłuszne.

Przypuśćmy teraz, że funkcja $\Psi(t)$ jest ujemna w przedziale $[0, T]$. Oznacza to, że sterowanie przyjmuje wartość -1 . Rozwiązaniem równania (6.25.2) jest

$$x(t) = -t + 1$$

natomiast równania (6.25.3)

$$\Psi(t) = 2 \left[-\frac{t^2}{2} + t + \frac{T^2}{2} - T \right]$$

Z wykresu funkcji (rys. 6.25C) widać, że jeżeli czas T jest mniejszy od jedności, to funkcja $\Psi(t)$ jest ujemna, a więc przypuszczenie jest słuszne. Pozostaje rozważyć przypadek, gdy T jest większe od jedności. Sterowanie może wówczas zmieniać znak w pewnej chwili z przedziału $[0, T]$.

Przyjmijmy, że w przedziale $[0, T]$ sterowanie jest ujemne (tzn. przyjmuje wartość -1), a zarazem ujemna jest funkcja $\Psi(t)$. Rezultaty poprzednich rozważań sugerują przyjęcie następującego rozwiązania:

$$u(t) = -1 \quad \text{dla } t \in [0, 1]$$

$$u(t) = 0 \quad \text{dla } t \in [1, T]$$

Wynika stąd rozwiązanie równań (rys. 6.25D)

$$x(t) = -t + 1 \quad \text{dla } t \in [0, 1]$$

$$x(t) = 0 \quad \text{dla } t \in [1, T]$$

$$\Psi(t) = 0 \quad \text{dla } t \in [1, T]$$

$$\Psi(t) = 2 \left(-\frac{t^2}{2} + t - \frac{1}{2} \right) \quad \text{dla } t \in [0, 1]$$

Widzimy, że rozwiązanie takie spełnia warunki konieczne wynikające z zasady maksimum.

Zastanówmy się nad interpretacją fizyczną tego zadania. Jak wiadomo, równaniem $\dot{x} = u$ opisuje się zbiornik cieczy przyłączony do pompy ssącej lub tłoczącej; u oznacza wydatek pompy, przy czym jeżeli $u = +1$, to pompa dostarcza wodę do zbiornika, a jeżeli $u = -1$ — to wodę pobiera. Nierówność

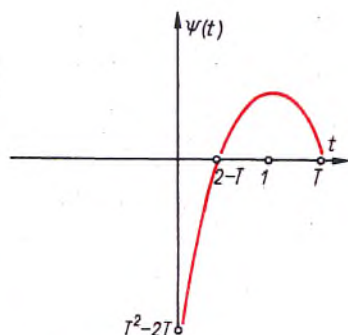
$$-1 \leq u \leq 1$$

oznacza, że określony jest maksymalny przepływ jaki może zapewnić pompa.

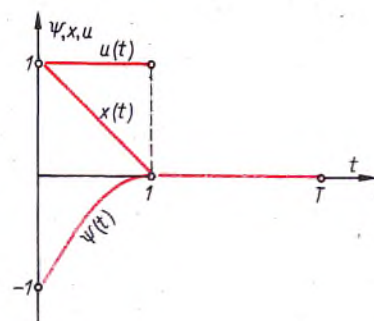
Funkcjonał jakości

$$J = \int_0^T x^2(t) dt$$

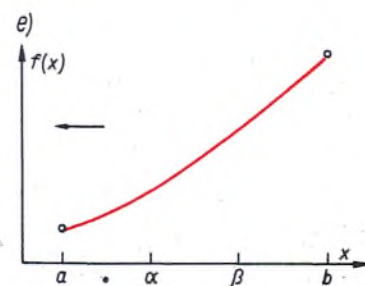
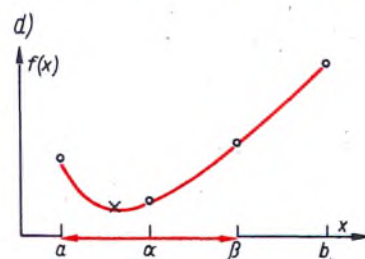
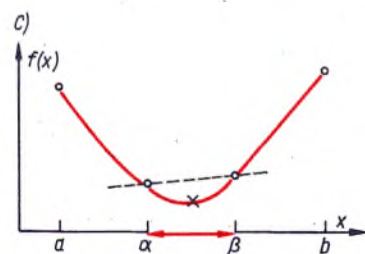
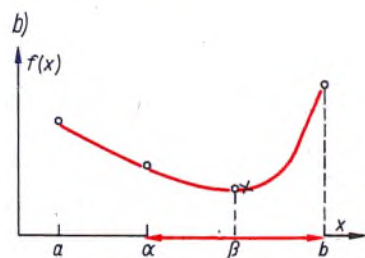
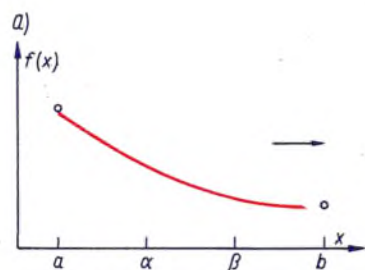
reprezentuje koszty eksploatacji zbiornika. Chwilowe koszty eksploatacji zbiornika są proporcjonalne do kwadratu poziomu cieczy. Zadanie, jakie stawiamy przed układem sterowania jest następujące:



•6.25C•



•6.25D•



Dobrać tak przepływ cieczy przez pompę, aby zapewnić minimalne koszty eksploatacji zbiornika.

Optymalne rozwiązanie jest w tym przypadku oczywiste — należy pobierać tak dużo wody ze zbiornika jak to jest możliwe (pompa pracuje na maksimum), po czym po opróżnieniu zbiornika pracę pompy należy przerwać. Zauważmy, że właśnie takie rozwiązanie otrzymaliśmy korzystając z zasady maksimum.

Niekiedy zadanie sterowania zbiornika formułuje się nieco inaczej — żąda się mianowicie, aby w chwili T w zbiorniku była określona pewna minimalna ilość cieczy. W ten sposób dochodzi się do sformułowania zadania optymalizacji z ograniczeniem

$$x(T) \geq x_{\min}$$

Rozwiązanie tego typu zadań za pomocą zasady maksimum jest również możliwe, jakkolwiek wymaga dużo większego wkładu obliczeń.

•6.26• W jaki sposób można wyznaczyć rozwiązanie rzeczywistego zadania optymalizacji?

Posługując się warunkami optymalności (zerowanie się pochodnych, warunki Kuhna-Tuckera) można efektywnie wyznaczyć (analityczne) rozwiązanie tylko dla bardzo prostych przypadków. Rzeczywiste zadania, zawierające wiele zmiennych decyzyjnych, ograniczeń oraz skomplikowaną funkcję celu prowadzą do układów równań nie dających się rozwikłać analitycznie. Należy w tym przypadku zastosować odpowiednie **metody numeryczne** poszukiwania ekstremum.

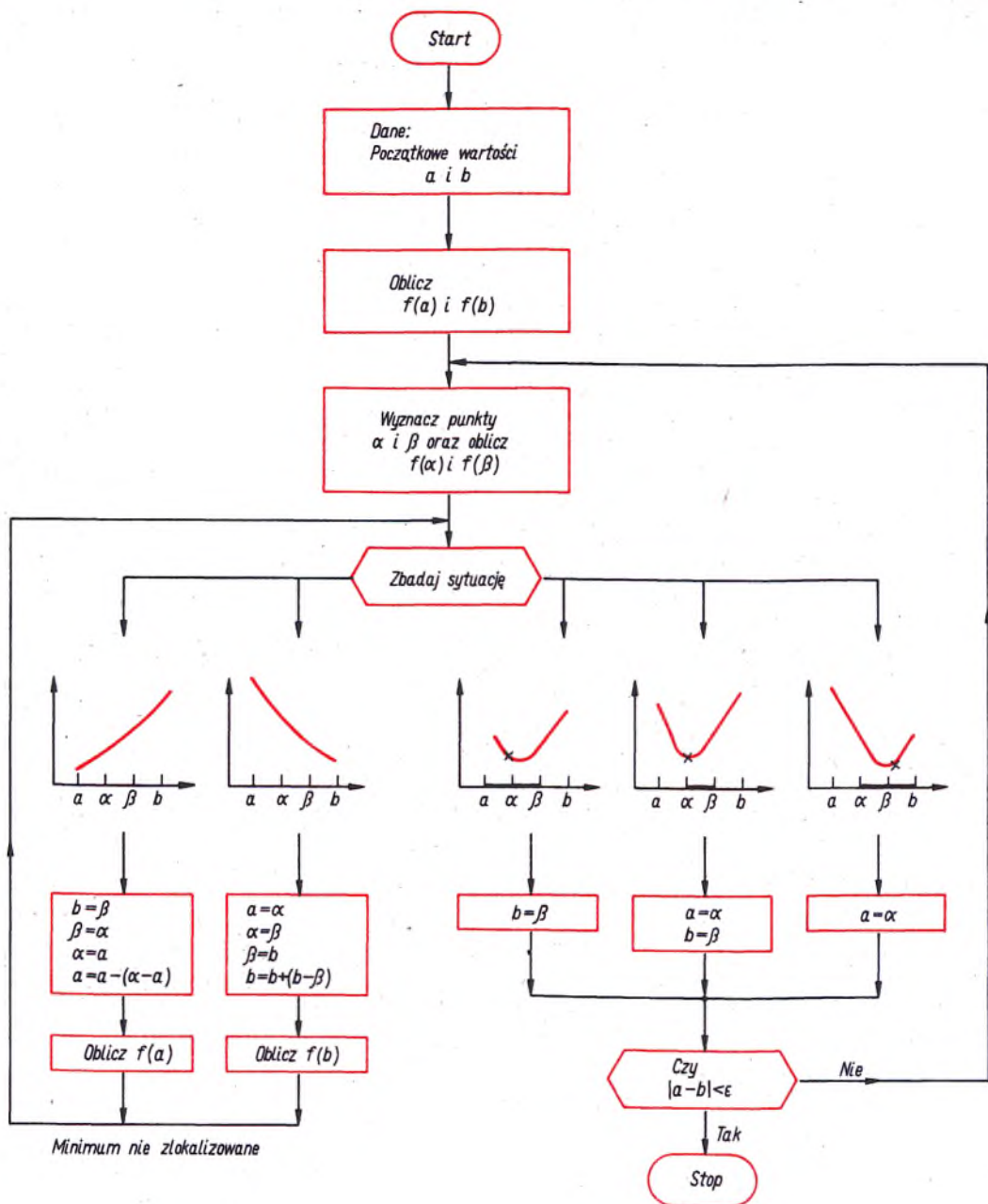
•6.27• W jaki sposób znaleźć numerycznie minimum funkcji jednej zmiennej?

Znanych jest wiele metod umożliwiających rozwiązanie tego zagadnienia. Podamy tylko jedną metodę, mimo prostoty charakteryzującą się dobrą efektywnością. Zakładamy, że jest znany przedział $[a, b]$, wewnątrz którego funkcja $f(x)$ posiada minimum. Zakładamy również, że jest to jedyne minimum w tym przedziale.

Działanie algorytmu jest następujące:

1. Odcinek $[a, b]$ dzielimy na 3 równe części punktami α i β i obliczamy wartości funkcji w tych punktach.
2. Na podstawie otrzymanych 4 wartości funkcji jesteśmy w stanie określić, który z trzech odcinków $[a, \beta]$, $[\alpha, \beta]$ zawiera minimum. Możliwe przypadki przedstawiono na rys. 6.27A.

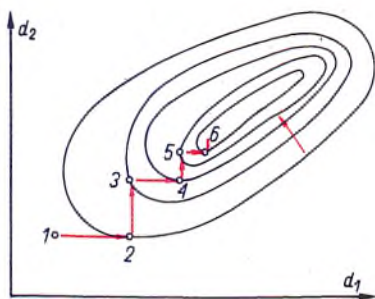
•6.27A• Możliwe sytuacje, jakie mogą wystąpić podczas minimalizacji funkcji jednej zmiennej. W przypadku (a) i (e) przedział $[a, b]$ nie zawiera minimum, konieczne jest poruszanie się w kierunku spadku funkcji (zaznaczono strzałką) w celu zlokalizowania przedziału



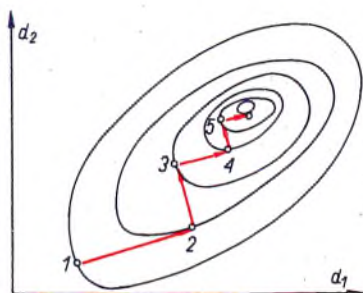
•6.27B• Schemat blokowy algorytmu poszukiwania minimum funkcji jednej zmiennej

Uwaga: na rysunku symbol „=” oznacza operację podstawienia, tzn. $\alpha = \beta$ oznacza, że pod zmienną α podstawia się wartość zmiennej β

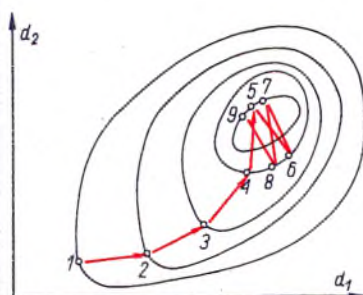
3. Po dokonaniu wyboru punkty końcowe odcinka zawierającego minimum traktujemy jako a i b i przechodzimy do punktu 1. Procedurę tę powtarzamy dopóty, dopóki długość odcinka zawierającego minimum nie jest dostatecznie mała. Działanie procedury przedstawiono dokładnie na rys. 6.27B.



•6.28A• Ilustracja metody Gaussa-Seidla



•6.28B• Ilustracja metody gradientowej minimalizacji funkcji z poszukiwaniem minimum w kierunku najszybszego spadku



•6.28C• Ilustracja metody gradientowej ze stałym krokiem. Zauważ, że działanie metody w pobliżu minimum

•6.28• W jaki sposób rozwiązać numerycznie zadanie optymalizacji bez ograniczeń w przypadku wielu zmiennych decyzyjnych?

Omówimy dwie najprostsze metody rozwiązywania tego typu zadań. Mimo prostoty charakteryzują się one dość dużą efektywnością, jednakże zawodzą w przypadku wielkiej liczby zmiennych decyzyjnych oraz w przypadku szczególnie „złośliwej” postaci funkcji celu.

a) Metoda Gaussa-Seidla

Działanie metod omówimy na przykładzie zadania z dwiema zmiennymi decyzyjnymi. Sposób ich zastosowania w bardziej złożonych przypadkach jest oczywisty. Chcąc zastosować metodę Gaussa-Seidla, wyróżnimy na płaszczyźnie (d_1, d_2) dwa prostopadłe kierunki. Wybieramy pewien punkt startowy (przybliżenie początkowe). Wyznaczamy minimum funkcji wzdłuż pierwszego kierunku.

Zauważmy że mamy do czynienia z zadaniem minimalizacji funkcji z **ograniczeniami równościowymi**, przy czym funkcja, którą minimalizujemy jest funkcją **jednej zmiennej** (zmiennej d_1 , bo zmienna d_2 pozostaje stała). Jak wiemy, chcąc znaleźć rozwiązanie takiego zadania, należy znaleźć linię poziomową funkcji odpowiadającą najmniejszej jej wartości, przy której linia ta jest styczna do prostej wzdłuż której szukamy minimum.

Zadanie to jest dość łatwe do rozwiązania numerycznego. Rozwiązując je otrzymamy nowy punkt (d_1, d_2); startując z tego punktu minimalizujemy funkcję wzdłuż **drugiego** kierunku. Otrzymany w wyniku tej minimalizacji punkt traktujemy jako punkt startowy i rozpoczynamy obliczenia od początku (rys. 6.28A).

b) metoda gradientu

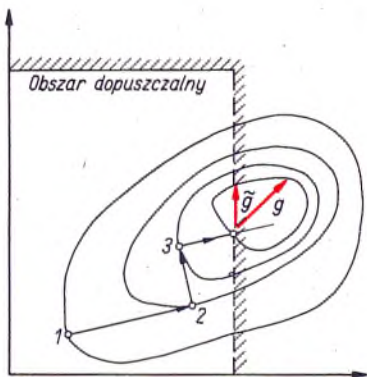
Wiadomo, że wektor gradientu wskazuje kierunek najszybszego wzrostu funkcji. Zatem, chcąc minimalizować funkcję, należy poruszać się w kierunku przeciwnym.

Załóżmy, że mamy dany punkt startowy (przybliżone początkowe rozwiązanie). Możemy obliczyć gradient w tym punkcie i wyznaczyć kierunek najszybszego spadku funkcji. Szukamy teraz minimum funkcji wzdłuż tego kierunku, a więc jak poprzednio mamy do czynienia z zadaniem minimalizacji funkcji **jednej zmiennej**. Zadanie to można łatwo rozwiązać. Otrzymany punkt traktujemy jako punkt startowy i powtarzamy obliczenia (rys. 6.28B).

Prostsza modyfikacja tej metody nie wymaga dokonywania minimalizacji wzdłuż kierunku najszybszego spadku, wystarczy wykonać tylko pewien krok w tym kierunku (rys. 6.28C). Oczywiście metoda taka może zawodzić przy niewłaściwym doborze kroku. Przedstawione metody zapewniają wyznaczenie rozwiązania w nieskończonej liczbie kroków. W praktyce oczywiście liczba kroków jest skończona —



•6.28D• Typowy przebieg wartości funkcji w kolejnych iteracjach algorytmu poszukiwania minimum. W przypadku złej metody można podjąć decyzję o zakończeniu obliczeń, mimo że znajdujemy się daleko od minimum



•6.29A• Metoda minimalizacji gradientowej w przypadku zadania z ograniczeniami nierównościowymi. Może się zdarzyć, że kierunek najszybszego spadku wyprowadza nas poza obszar dopuszczalny. Musimy poruszać się wówczas wzdłuż kierunku nie wyprowadzającego ze zbioru dopuszczalnego i najbliższego kierunkowi najszybszego spadku

działanie procedury przerywa się w chwili, gdy np. różnica między wartościami funkcji w dwóch kolejnych wyznaczonych punktach jest dostatecznie mała (funkcja zaczyna maleć bardzo wolno) (rys. 6.28D).

•6.29• W jaki sposób rozwiązać numerycznie zadanie optymalizacji z ograniczeniami w przypadku wielu zmiennych decyzyjnych?

Na przykładzie zadania z dwoma zmiennymi decyzyjnymi przedstawimy dwie najczęściej spotykane metody rozwiązywania zadań z ograniczeniami — metodę rzutu gradientu i metodę funkcji kary.

a) Metoda rzutu gradientu

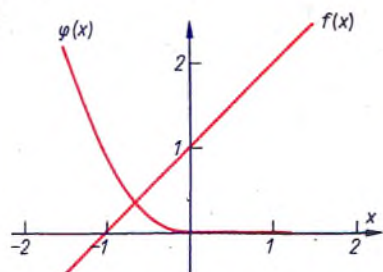
Jak wiadomo z poprzedniego pytania, kierunek przeciwny do kierunku wektora gradientu określa kierunek najszybszego spadku funkcji. Chcąc rozwiązać zadanie optymalizacji należy minimalizować funkcję wzdłuż tego kierunku. Może się jednak zdarzyć, że rozwiązanie zadania minimalizacji funkcji jednej zmiennej wyprowadza nas poza zbiór decyzji dopuszczalnych (rys. 6.29A). Sposób postępowania jest wówczas następujący: Poruszamy się wzdłuż wyznaczonego kierunku najszybszego spadku tak daleko, jak to jest tylko możliwe bez naruszania ograniczeń, tzn. aż do brzegu obszaru dopuszczalnego. Wyznaczony punkt traktujemy jako przybliżenie startowe i wyznaczamy gradient w tym punkcie. Jeżeli kierunek najszybszego spadku prowadzi do wnętrza obszaru dopuszczalnego, wykonujemy kolejny krok metody gradientu (tzn. minimalizujemy funkcję wzdłuż tego kierunku). W przeciwnym przypadku staramy się znaleźć kierunek „najbliższy” kierunkowi najszybszego spadku, nie wyprowadzający jednak poza zbiór. Oczywiście kierunek taki jest równoległy do krawędzi ograniczeń. Mając wyznaczony kierunek można rozwiązać zadanie minimalizacji funkcji jednej zmiennej. Przykład działania procedury pokazano na rys. 6.29A.

b) Metoda funkcji kary

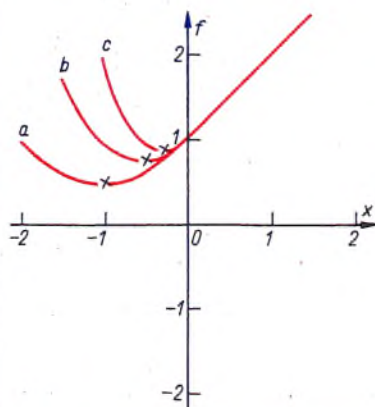
Metoda ta polega na takiej modyfikacji funkcji celu, aby przyjmowała ona bardzo duże wartości poza obszarem dopuszczalnym; do minimalizacji takiej funkcji stosujemy jedną z opisanych poprzednio procedur minimalizacji bez ograniczeń. W ten sposób procedura minimalizacji jest niejako „karana” za przekroczenie ograniczeń („nie opłaca” się jej wchodzić zbyt daleko poza obszar dopuszczalny). Przykład. Znaleźć minimum funkcji $f(x) = x+1$ przy ograniczeniu $x \geq 0$.

Łatwo podać sposób modyfikacji funkcji tak, aby jej wartość rosła bardzo szybko przy przekroczeniu ograniczeń. Funkcja zmodyfikowana ma postać

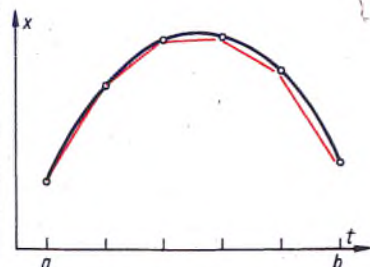
$$f(x) = x+1 + \varrho \varphi(x)$$



• 6.29B • Funkcja minimalizowana $f(x)$ oraz funkcja kary $\varphi(x)$



• 6.29C • Ilustracja działania metody funkcji kary: a) funkcja zmodyfikowana dla $\rho = \frac{1}{2}$; b) dla $\rho = 1$; c) dla $\rho = 2$



• 6.30 • Sposób przybliżania funkcji ciągłej funkcją odcinkowo-liniową

$$\text{gdzie } \varphi(x) = \begin{cases} x^2 & \text{dla } x < 0 \\ 0 & \text{dla } x \geq 0 \end{cases}$$

a ρ jest dużą liczbą dodatnią (rys. 6.29B). Szukamy minimum funkcji $f(x)$

$$\frac{\partial f}{\partial x} = 1 + \rho \varphi'(x)$$

gdzie

$$\varphi'(x) = \begin{cases} 2x & \text{dla } x < 0 \\ 0 & \text{dla } x \geq 0 \end{cases}$$

Pochodna zeruje się dla $x = -\frac{1}{2\rho}$; łatwo sprawdzić, że tylko w tym punkcie funkcja zmodyfikowana ma minimum. Zauważmy, że dobierając dostatecznie duże ρ otrzymamy rozwiązanie zadania zmodyfikowanego różniące się dowolnie mało od rozwiązania zadania pierwotnego (rys. 6.29C).

• 6.30 • W jaki sposób można rozwiązać numerycznie zadanie wariacyjne?

Najprostszym sposobem rozwiązania takiego zadania jest sprowadzenie go do zadania minimalizacji funkcji wielu zmiennych. Załóżmy, że poszukujemy funkcji $x(t)$ określonej w przedziale $[0, T]$, zapewniającej minimum pewnego funkcjonału. Przedział $[0, T]$ dzielimy na pewną liczbę równych odcinków.

Założmy, że znamy wartości funkcji $x(t)$ w punktach podziału. Łącząc odcinkami punkty odpowiadające wartościom funkcji otrzymamy pewne **przybliżenie** funkcji $x(t)$, w postaci funkcji odcinkowo-liniowej (łamej) (rys. 6.30).

Zauważmy, że funkcję taką można jednoznacznie określić podając jej wartości w skończonej liczbie punktów. Formułujemy teraz zadanie **przybliżone** — znaleźć odcinkowo-liniową funkcję $x(t)$ zapewniającą minimum funkcjonału jakości. Po podstawieniu takiej funkcji do funkcjonału okaże się, że zależy on tylko od skończonej liczby zmiennych — będą to wartości funkcji w punktach podziału. Tak więc zadanie minimalizacji funkcjonału sprowadziliśmy do zadania minimalizacji funkcji wielu zmiennych. Taki sposób postępowania nosi nazwę **metody Ritz**.

• 6.31 • Jakie problemy automatyki prowadzą do zadań optymalizacji?

Rozważane dotychczas problemy teoretyczne optymalizacji i metody obliczeniowe znajdują zastosowanie w bardzo wielu dziedzinach życia. Można tu wymienić np. problemy optymalnych inwestycji w przemyśle, problemy optymalizacji konstrukcji inżynierskich (np. budowa mostów), problemy optymalnego składu pasz w hodowli itp.

W ciągu ostatnich kilkunastu lat nastąpiło znaczne przeniknięcie metod matematycznych (w tym i optymalizacji) do

automatyki. Stało się to możliwe dzięki rozwojowi zarówno samej teorii sterowania, jak i elektronicznej techniki obliczeniowej.

Wymienimy najważniejsze dziedziny zastosowań teorii optymalizacji w automatyce. Są to:

- optymalizacja statyczna
- optymalizacja parametryczna
- problem regulatora optymalnego
- problem identyfikacji modeli matematycznych
- regulacja ekstremalna
- adaptacyjne układy sterowania.

•6.32• Co to jest optymalizacja statyczna?

Jak wiadomo, obiekty którymi interesuje się automatyka są na ogół obiektami dynamicznymi, tzn. takimi, których zachowanie się w przyszłości zależy między innymi od tego, jakie sygnały działały na ten obiekt w przeszłości. Jeżeli taki obiekt jest stabilny, to po dostatecznie długim czasie wszystkie współrzędne (zmienne stanu) ustalają się. Oczywiście fakt ten może mieć miejsce tylko wówczas, gdy wymuszenia działające na obiekt są stałe w czasie lub zmieniają się bardzo powoli. Zadaniem regulatora jest utrzymanie określonego stanu ustalonego. Sygnałem wejściowym regulatora jest między innymi wartość zadana wielkości regulowanej. Wartość ta może być ustalana np. przez technologa projektującego proces.

Wybierając wartości zadane dla regulatorów technolog nie jest w stanie uwzględnić wszystkich czynników działających na proces. Z kolei od wyboru wartości zadanych dla regulatorów (a więc punktów pracy poszczególnych elementów systemu) zależy w sposób istotny wydajność instalacji.

Powstaje następujący problem: **w jaki sposób wybrać najlepszą wartość zadaną dla regulatora?**

Zadanie wyboru stanu ustalonego można sformułować w postaci odpowiedniego problemu optymalizacji; zadanie to nazywamy **optymalizacją statyczną układu** lub **optymalizacją stanu ustalonego**.

•6.33• Jak jest typowe zadanie optymalizacji statycznej?

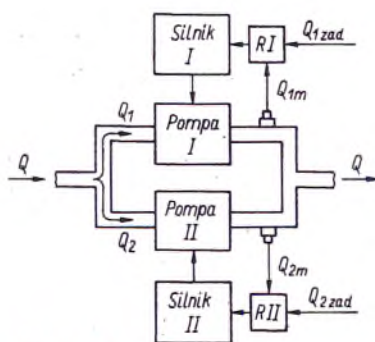
Rozważania przeprowadzimy na przykładzie. W przepompowni pracują dwie pompy połączone równolegle. Każda z pomp jest wyposażona w regulator przepływu. Regulator ten tak zmienia prędkość obrotową silnika napędzającego pompę, aby uzyskać żądany przepływ (rys. 6.33A).

Jeżeli przez Q_1 i Q_2 oznaczymy przepływ przez pierwszą i drugą pompę, to koszt eksploatacji pomp będzie następujący:

koszt eksploatacji pompy I — $2Q_1^2$

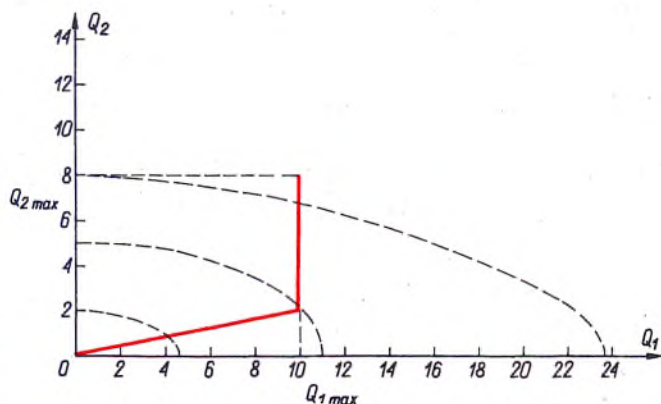
koszt eksploatacji pompy II — $10Q_2^2$

Jednocześnie maksymalne przepływy możliwe do zrealizowania przez poszczególne pompy są następujące:



•6.33A• Schemat przepompowni wraz z układami regulacji

•6.33B• Poziomice wskaźnika jakości, zbiór sterowań dopuszczalnych oraz zbiór rozwiązań optymalnych zadania optymalizacji statycznej przepompowni



pompa I — $Q_{max} = 10$

pompa II — $Q_{max} = 8$

Jak dobrać wartości zadane regulatorów obu pomp aby sumaryczny przepływ był równy Q ? Przeanalizujemy odpowiedź na to pytanie. Jeżeli przepływ przez pierwszą pompę jest równy Q_1 , to przepływ przez drugą pompę jest równy $Q - Q_1$. Wynika stąd, że Q_1 można wybrać **dowolnie**. Zmieniając wartość Q_1 otrzymamy różne koszty eksploatacji całej instalacji. Można więc wybrać Q_1 tak, aby uzyskać koszty jak najmniejsze.

Tak więc problem doboru wartości zadanych dla regulatorów pomp można sformułować następująco:

Dobrac przepływy Q_1 i Q_2 spełniające nierówności

$$0 \leq Q_1 \leq 10$$

$$0 \leq Q_2 \leq 8$$

$$Q_1 + Q_2 = Q$$

i zapewniające minimum funkcji

$$f(Q_1, Q_2) = 2Q_1^2 + 10Q_2^2$$

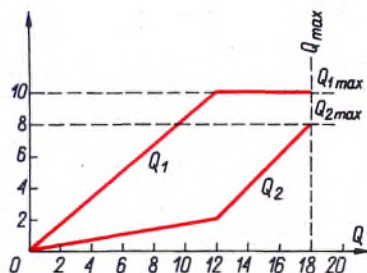
Jest to typowe zadanie optymalizacji. Zadanie to można łatwo rozwiązać korzystając z jego interpretacji graficznej. Na rysunku 6.33B przedstawiono zbiór decyzji dopuszczalnych i poziomice funkcji celu. W wyniku analizy rysunku otrzymujemy następujące rozwiązanie (rys. 6.33C):

gdy $Q > 18$ — rozwiązania nie ma

$$\text{gdy } 0 \leq Q \leq 12 \quad Q_2 = \frac{1}{6}Q, \quad Q_1 = \frac{5}{6}Q$$

$$\text{gdy } 12 \leq Q \leq 18 \quad Q_1 = 10, \quad Q_2 = Q - 10$$

Otrzymaliśmy zatem wzory umożliwiające obsłudze przepompowni nastawianie wartości zadanych dla regulatorów w zależności od żądanego całkowitego przepływu przez przepompownię.



•6.33C• Zależność optymalnych wartości przepływu przez poszczególne pompy od sumarycznego przepływu przez przepompownię

•6.34• Co to jest optymalizacja parametryczna układu regulacji?

Na etapie konstruowania układu regulacji automatycznej powstaje problem doboru nastaw regulatora (np. wzmocnienia, czasu całkowania itp). Istnieje cały szereg metod umożliwiających rozwiązanie tego typu zadań. Metody te nie dają na ogół jednoznacznej odpowiedzi. Projektant układu regulacji może wybierać parametry regulatora różnymi sposobami, otrzymując za każdym razem nieco inne właściwości układu regulacji. Ostateczny dobór parametrów przeprowadza się na podstawie założeń projektowych oraz intuicji i doświadczenia projektanta.

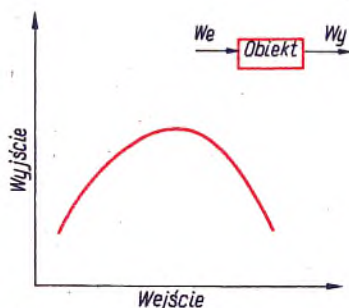
Projektując układ regulacji można postąpić w nieco inny sposób. Można mianowicie ilościowo scharakteryzować pewne właściwości układu regulacji. Ilościową miarą jakości regulacji może być np. tłumienie zakłóceń, szerokość pasma i inne wskaźniki. Oczywiście wartość tych wskaźników zależy od doboru parametrów regulatora. Mamy zatem określony pewien wskaźnik jakości charakteryzujący jakość projektowanego układu; zadaniem projektanta jest taki dobór parametrów, aby uzyskać maksimum wskaźnika charakteryzującego jakość działania układu.

Zadanie projektowania układu regulacji sprowadzono zatem do zadania optymalizacji; zmienną decyzyjną są w tym zadaniu parametry regulatora. Stąd też nazwa tego typu podejścia — **optymalizacja parametryczna**.

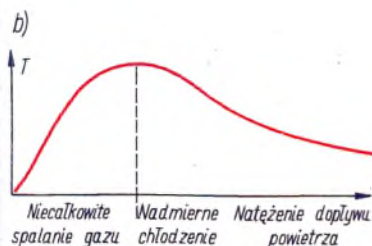
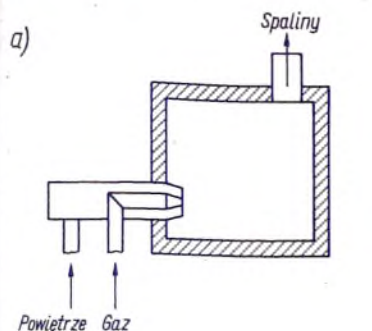
Podkreślić należy, że wybór odpowiedniego wskaźnika jakości jest bardzo trudny, a otrzymane rezultaty zależą w istotny sposób od jego postaci. Tak więc zadanie syntezy układu regulacji nie zostało wbrew pozorom uproszczone — zamiast dobierać bezpośrednio parametry regulatora projektant musi dobierać wskaźnik jakości. Taki sposób postępowania w niektórych przypadkach może okazać się celowy — dotyczy to zwłaszcza układów o dużej liczbie parametrów, których bezpośredni dobór jest trudny.

•6.35• Co to jest regulacja ekstremalna?

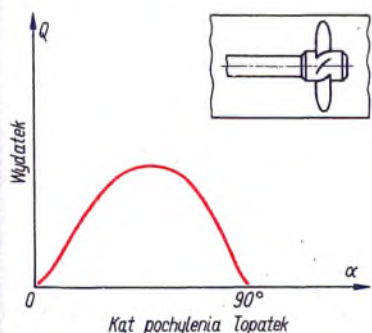
Założmy, że mamy do czynienia z pewnym obiektem statycznym, którego charakterystyka ma kształt pokazany na rys. 6.35. Obiekty mające taką charakterystykę nazywamy **objektami ekstremalnymi**. Zadaniem układu regulacji ekstremalnej jest takie dobranie sterowania (sygnału wejściowego) aby utrzymać sygnał **wyjściowy na jak najwyższym poziomie**. W pierwszej chwili zadanie wydaje się łatwe do rozwiązania — wystarczy raz wyznaczyć sterowanie zapewniające maksymalizację sygnału wyjściowego obiektu, a potem sterowanie takie zastosować. Jednak w praktyce sytuacja jest znacznie bardziej złożona — charakterystyka obiektu nie jest stała, lecz zmienia się wskutek zakłóceń działających na obiekt. Zakłócenia te są na ogół niemierzalne (lub trudno mierzalne), a ponadto nie wiadomo, jak wpływają na cha-



•6.35• Obiekt o charakterystyce ekstremalnej



•6.36A• Schemat budowy pieca gazowego i jego charakterystyka statyczna



•6.36B• Charakterystyka statyczna wentylatora z regulowanym kątem ustawienia łopatek

rakterystykę (położenie ekstremum). Z tych względów opracowano regulatory, które samoczynnie dobierają sygnał wejściowy tak, aby uzyskać maksimum sygnału wyjściowego obiektu. Regulatory takie nazywamy **regulatorami ekstremalnymi**.

•6.36• Jak można podać przykłady obiektów ekstremalnych?

Typowym przykładem obiektu ekstremalnego jest piec zasilany gazem. Jest to w istocie obiekt dynamiczny, jednakże przy założeniu, że sygnały wejściowe zmieniają się dostatecznie wolno rozważać możemy tylko jego charakterystykę statyczną.

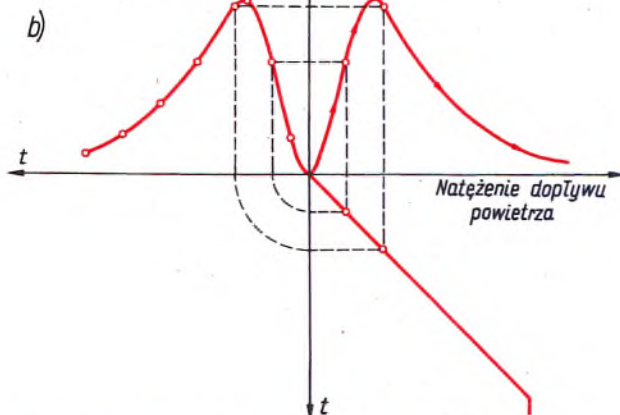
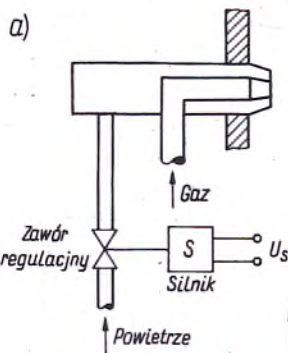
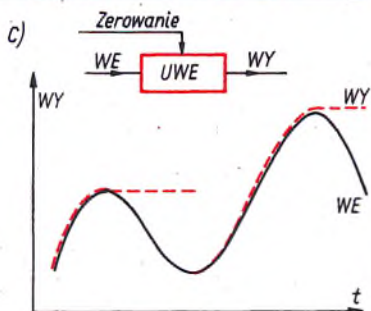
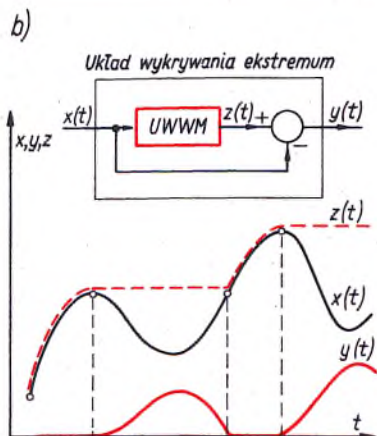
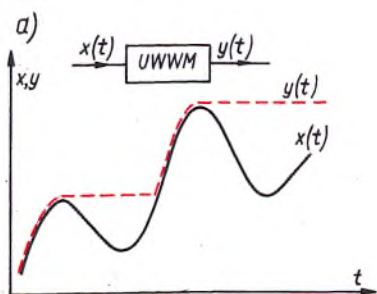
Na rysunku 6.36A(a) przedstawiono schemat pieca. Do palnika doprowadzony jest osobno gaz i powietrze. Załóżmy, że natężenie dopływu gazu jest stałe. Zbadajmy wpływ zmiany natężenia dopływu powietrza na temperaturę pieca. Jeżeli dopływ powietrza jest równy zeru, spalanie gazu nie zachodzi z powodu braku tlenu. Jeżeli natomiast dopływ powietrza jest bardzo duży, to część tlenu jest zużyta do spalania gazu, natomiast pozostała część chłodzi wnętrze pieca. Chłodzenie to jest tym intensywniejsze, im większy jest dopływ powietrza. Z powyższych rozważań wynika, że charakterystyka statyczna pieca jest charakterystyką typu ekstremalnego (rys. 6.36A(b)).

Dokładny przebieg charakterystyki jest jednak nieznamy — zależy on między innymi od wartości opałowej gazu, właściwości wsadu znajdującego się wewnątrz pieca itp. Zauważmy ponadto, że kształt charakterystyki może ulegać zmianom w czasie. Zmiany mogą być spowodowane np. wyparowywaniem wody ze wsadu co powoduje zmianę ciepła właściwego materiału.

Innym przykładem obiektu ekstremalnego jest wentylator z regulowanym kątem ustawienia łopatek (rys. 6.36B). Jeżeli kąt ustawienia łopatek jest równy 0° , lub 90° , to przepływ powietrza jest równy zeru. Istnieje zatem takie ustawienie, przy którym wydatek wentylatora jest maksymalny.

•6.37• Jak działa regulator ekstremalny?

Zasadę działania regulatora ekstremalnego wyjaśnimy na przykładzie układu regulacji temperatury pieca. Załóżmy, że regulacja dopływu powietrza następuje za pośrednictwem zaworu napędzanego silnikiem elektrycznym (rys. 6.37A(a)). Oznacza to, że po przyłożeniu stałego napięcia na zaciski silnika natężenie dopływu powietrza zmienia się liniowo w czasie. Na rysunku 6.37A(b) przedstawiono przebieg natężenia dopływu powietrza i temperatury pieca. Załóżmy teraz, że sygnał proporcjonalny do temperatury wewnątrz pieca (np. napięcie z termoelementu) podajemy na **układ wybierania wartości maksymalnej**. Układ ten jest prostym urządzeniem elektronicznym zapamiętującym największą

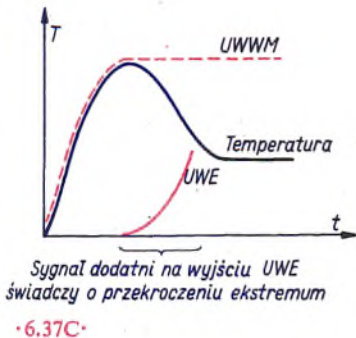


•6.37A• Regulacja temperatury w piecu gazowym przy liniowej zmianie dopływu powietrza: a) realizacja; b) sposób konstruowania funkcji $T(t)$ na podstawie znajomości charakterystyki ekstremalnej pieca. Strzałkami zaznaczono ruch punktu pracy na charakterystyce statycznej obiektu

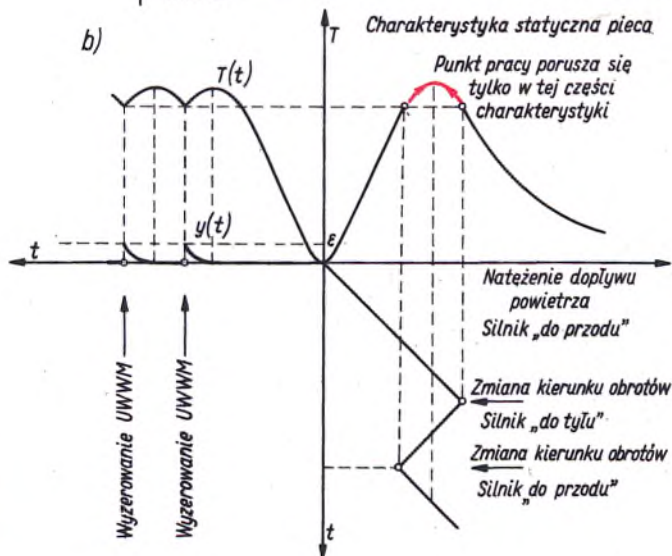
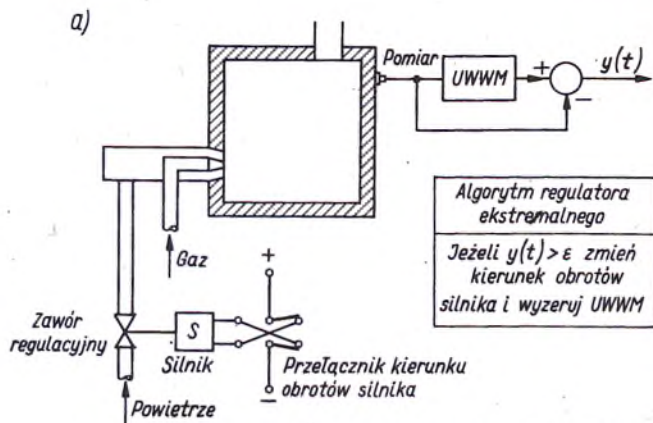
wartość sygnału wejściowego. Działanie takiego urządzenia pokazano na rys. 6.37B. Jeżeli od sygnału wyjściowego urządzenia wybierania wartości maksymalnej odejmiemy sygnał proporcjonalny do temperatury, to otrzymamy w ten sposób **układ wykrywania ekstremum**. Pracę takiego urządzenia ilustruje rysunek.

Załóżmy, że na zaciski silnika regulującego dopływ powietrza podaliśmy napięcie stałe o takiej polaryzacji, że dopływ powietrza wzrasta. Poruszamy się więc po rosnącej części charakterystyki statycznej. Ponieważ temperatura wzrasta to sygnał na wyjściu urządzenia wybierania wartości ekstremalnej jest taki sam, jak na wejściu. Wynika stąd, że na wyjściu układu wykrywania ekstremum jest zero.

•6.37B• Działania układu wykrywania ekstremum: a) zasada działania układu wybierania wartości maksymalnej (UWWM); b) sposób wykorzystania układu wybierania wartości maksymalnej do budowy układu wykrywania ekstremum; c) działanie układu wykrywania ekstremum (UWE) z zerowaniem



•6.37C•



•6.37D• Działanie ekstremalnego regulatora temperatury z układem wykrywania ekstremum (a), oraz przebiegi w układzie regulacji (b)

Jeżeli punkt pracy pieca (dopływ powietrza) przekroczy ekstremum, na wyjściu urządzenia do wybierania wartości maksymalnej pojawi się stały sygnał równy maksymalnej temperaturze możliwej do uzyskania w danych warunkach (rys. 6.37C). Na wyjściu urządzenia do wykrywania ekstremum pojawi się zatem pewien sygnał ujemny (temperatura pieca zaczyna maleć po przekroczeniu ekstremum). Pojawienie się takiego sygnału świadczy o konieczności zmiany polaryzacji napięcia zasilającego silnik. Jeżeli dokonamy tej czynności, a zarazem wyzerujemy układ wybierania wartości maksymalnej, to cały cykl powtórzy się — dopływ powietrza będzie zmniejszał się tak długo, aż na wyjściu układu wykrywania ekstremum pojawi się sygnał ujemny. Należy wówczas zmienić polaryzację napięcia zasilającego silnik. Czynności te łatwo zautomatyzować stosując zwykły przekaźnik ze strefą nieczułości. Otrzymuje się w ten sposób **zamknięty układ regulacji ekstremalnej** (rys. 6.37D). Zasadniczą cechą

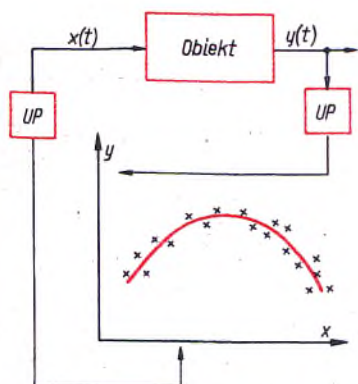
takiego układu jest istnienie w nim **oscyłacji niegasnących**. Dzięki tym oscylacjom układ przeszukuje otoczenie ekstremum; jeżeli charakterystyka układu przesuwa się z jakiś powodów, układ nadąża za przesuującym się ekstremum.

Należy zaznaczyć, że poszukiwanie ekstremum odbywa się **bezpośrednio na obiekcie**. Z tego względu zastosowanie regulatorów ekstremalnych jest mocno ograniczone. Nie można ich stosować np. do regulacji procesów chemicznych, gdyż znaczne odchylenia od pożądanego punktu pracy mogą być bardzo szkodliwe (np. mogą przerwać reakcję lub spowodować wybuch reaktora). Również zastosowanie takich regulatorów do obiektów o bardzo dużej stałej czasowej jest niecelowe, gdyż wymaga zastosowania oscylacji przeszukujących o bardzo małej częstotliwości. W takich przypadkach są stosowane **regulatory adaptacyjne z modelem procesu**.

•6.38• Co to jest układ adaptacyjny regulacji ekstremalnej?

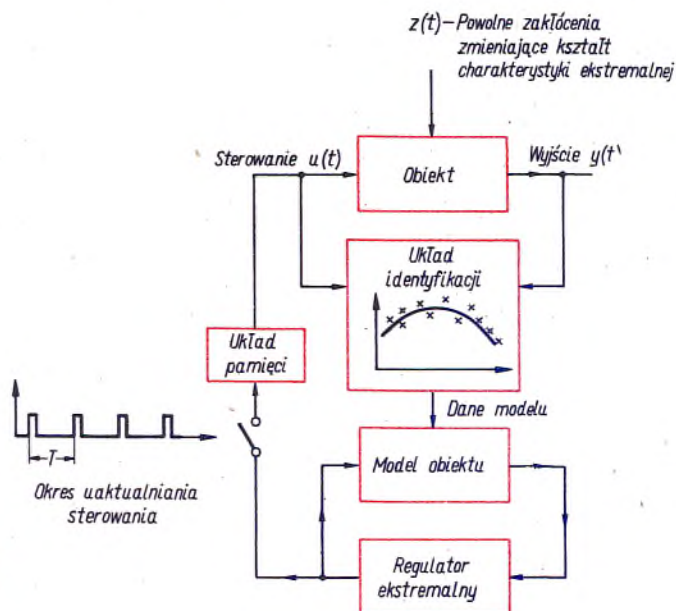
Załóżmy, że mamy do czynienia z obiektem o charakterystyce ekstremalnej. Dla uproszczenia rozważań założymy, że jest to obiekt statyczny.

Podając sygnały (sterowania) na wejście obiektu, na jego wyjściu obserwować możemy sygnały wyjściowe (odpowiedzi). Nanosząc te wartości na wykres można wyznaczyć charakterystykę statyczną obiektu (rys. 6.38A). Proces ten nazywamy **identyfikacją** charakterystyki statycznej obiektu. Jeżeli dokonaliśmy takiej identyfikacji, to możemy wyznaczyć ekstremum charakterystyki statycznej i sterowanie zapewniające utrzymanie punktu pracy obiektu w punkcie экстре-



•6.38A• Sposób identyfikacji charakterystyki statycznej obiektu

UP — układ pomiarowy



•6.38B• Schemat blokowy adaptacyjnego układu regulacji ekstremalnej z układem identyfikacji obiektu i przeszukiwaniem na modelu

malnym oraz sterowanie to przyłożyć do obiektu. Gdyby charakterystyka statyczna nie zmieniała się pod wpływem zakłóceń, operację taką należałoby wykonać tylko raz. W praktyce model matematyczny obiektu należy co jakiś czas uaktualniać. W tym celu przez cały czas pracy układu mierzone są sygnały wejściowe i wyjściowe obiektu. Po zebraniu dostatecznie dużej liczby danych wykonuje się powtórzną identyfikację modelu i wyznacza się sterowanie. Taki sposób postępowania zapewnia utrzymanie maksymalnego sygnału wyjściowego obiektu, przy czym poszukiwanie tego ekstremum jest wykonywane **na modelu** (rys. 6.38B). Operację tę można wykonywać bardzo szybko. W praktyce, ze względu na złożoność zadań wykonywanych przez regulator układu takie realizowane są za pomocą maszyn matematycznych cyfrowych i analogowych.

•6.39• Co to jest identyfikacja modelu matematycznego obiektu?

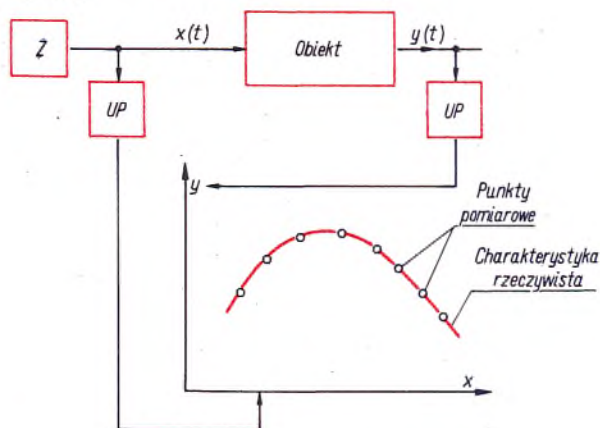
Jak wiadomo, podstawowym problemem przy konstruowaniu układu regulacji automatycznej jest sformułowanie modelu matematycznego obiektu. Model ten, w postaci układu równań formułuje się w wyniku rozważań teoretycznych, na podstawie bilansu wielkości fizycznych. W tak określonych równaniach występuje cały szereg współczynników liczbowych charakteryzujących proces. Współczynniki te mogą zależeć np. od rozmiarów reaktora chemicznego, właściwości produktów itp. O ile rozmiary reaktora łatwo jest określić, o tyle inne parametry związane np. z ciepłem właściwym i innymi właściwościami substancji reagujących są trudne do określenia. Pozostaje więc eksperymentalne określenie tych współczynników.

Eksperymentalne określenie parametrów modelu matematycznego obiektu nazywamy **identyfikacją parametryczną**. Czasem zadanie identyfikacji jest stawiane znacznie ogólniej — na podstawie danych doświadczalnych należy określić zarówno model matematyczny jak i jego parametry. Sytuacja taka może zachodzić wówczas, gdy ze względu na znaczną złożoność obiektu nie jesteśmy w stanie przeprowadzić analizy teoretycznej zjawisk. Mówi się wówczas, że obiekt jest traktowany jako „czarna skrzynka”.

•6.40• W jaki sposób dokonuje się identyfikacji obiektu statycznego?

Założmy, że identyfikowany obiekt jest wyposażony w mierniki sygnałów wejściowych i wyjściowych. Dla uproszczenia rozważać będziemy obiekt o jednym wejściu i jednym wyjściu. Obserwując działanie obiektu w ciągu dostatecznie długiego czasu można zarejestrować wartości sygnałów działających na obiekt i odpowiadających im sygnałów wyjściowych. Nanosząc te wartości na płaszczyznę z odpowiednim układem współrzędnych, teoretycznie powinniśmy otrzymać charakterystykę statyczną obiektu (rys. 6.40A).

Nadajnik sygnału
wejściowego



•6.40A• Identyfikacja obiektu statycznego w przypadku braku zakłóceń pomiarowych i zakłóceń działających na obiekt

W praktyce problem jest bardziej skomplikowany. Na obiekt działają nieustannie zakłócenia zmieniając nieco kształt charakterystyki. Zakłóceń tych nie można na ogół zmierzyć. Ponadto pomiary sygnałów wejściowych i wyjściowych są dokonywane z pewnym błędem. Czynniki te powodują, że punkty pomiarowe naniesione na płaszczyznę układają się w „chmurkę” i na pierwszy rzut oka trudno powiedzieć coś o dokładnym kształcie charakterystyki statycznej (rys. 6.40B(a)). Przypuśćmy, że na podstawie analizy teoretycznej działania obiektu przewidujemy, że charakterystyka statyczna ma postać funkcji

$$y = a_0 + a_1 x + a_2 x^2$$

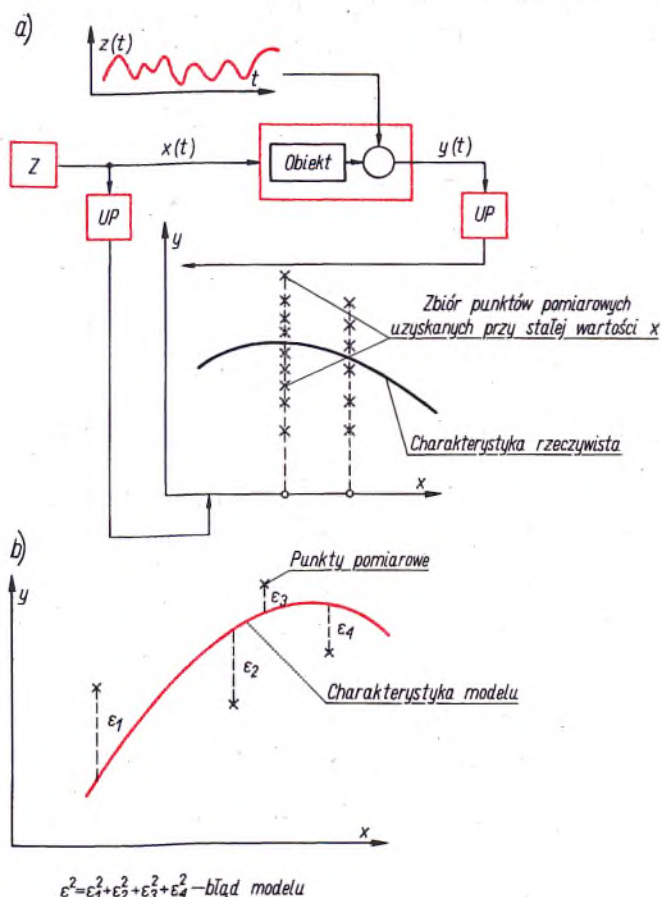
Problem identyfikacji sformułować można teraz następująco: na podstawie mierzonych wartości x i y określić współczynniki a_0 , a_1 i a_2 .

Gdyby pomiary były wykonywane absolutnie dokładnie i na obiekt nie działały żadne zakłócenia, to do określenia tych współczynników wystarczyłyby wyniki tylko 3 pomiarów. W praktyce, ze względu na znaczny rozrzut wartości pomiarowych poszukuje się współczynników funkcji tak, aby ta funkcja w pewnym sensie **najlepiej przybliżała** zbiór punktów pomiarowych. Pozostaje zatem sprecyzować sens najlepszego przybliżenia.

Rozważmy wyniki i -tego pomiaru. Wartość sygnału wejściowego jest równa x_i , wyjściowego — y_i . Gdyby wykres funkcji przechodził przez punkt o współrzędnych (x_i, y_i) , to miałyby miejsce równość

$$\varepsilon_i = y_i - a_0 - a_1 x_i - a_2 x_i^2 = 0$$

Jeżeli punkt nie leży na wykresie funkcji, to wyrażenie powyższe nie jest równe zero. Bezwzględna wartość tego wyrażenia jest w pewnym sensie odległością punktu od wykresu funkcji. Obliczymy teraz sumę odległości w s z y s t k i c h



6.40B Identyfikacja obiektu statycznego w warunkach rzeczywistych:
a) na obiekt działają zakłócenia $z(t)$ uniemożliwiające dokładny pomiar charakterystyk; b) sposób wyznaczania błędu modelu

punktów pomiarowych od wykresu funkcji (rys. 6.40B(b)). Suma ta zależy oczywiście od wartości współczynników a_0, a_1 i a_2 . Sformułujemy zadanie optymalizacji: dobrać współczynniki a_0, a_1 i a_2 tak, aby zminimalizować sumę odległości wszystkich punktów pomiarowych od wykresu funkcji. Formalnie zadanie to zapisujemy

$$\min_{a_0, a_1, a_2} \sum_i |\epsilon_i|$$

Ponieważ funkcja ta jest trudna do minimalizacji, w praktyce stosuje się nieco inną jej postać

$$\min_{a_0, a_1, a_2} \sum_i \epsilon_i^2$$

Okazuje się, że taka funkcja ma pewne dodatkowe właściwości — eliminuje w pewnym sensie wpływ zakłóceń pomiarowych.

Minimalizacja takiej funkcji jest zadaniem stosunkowo prostym. Można zastosować jakiegokolwiek algorytm minima-

lizacji numerycznej lub skorzystać z warunków koniecznych optymalności.

•6.41• W jaki sposób dokonuje się identyfikacji obiektu dynamicznego?

Zasada identyfikacji obiektu dynamicznego jest zbliżona do zasady identyfikacji obiektu statycznego.

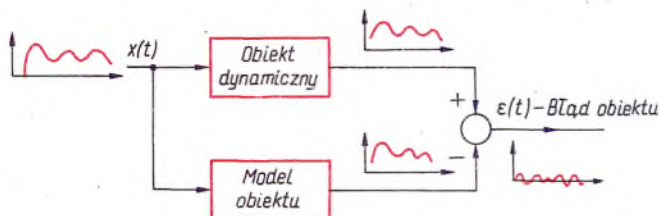
Załóżmy, że na obiekt dynamiczny działają wymuszenia — układ znajduje się w warunkach normalnej eksploatacji. Można więc zarejestrować sygnały działające na obiekt i odpowiadające im sygnały wyjściowe. Jeśli zarejestrowane sygnały wejściowe przyłożymy do modelu obiektu (rozwiążemy równania modelu), to na ogół okaże się, że rozwiązania uzyskane z modelu różnią się od zarejestrowanych rzeczywistych sygnałów wyjściowych (rys. 6.41A). Główną przyczyną różnicy jest niezgodność parametrów modelu i obiektu. Miarą niezgodności może być całka kwadratu błędu

$$J = \int_0^T (\hat{y}(t) - y(t))^2 dt$$

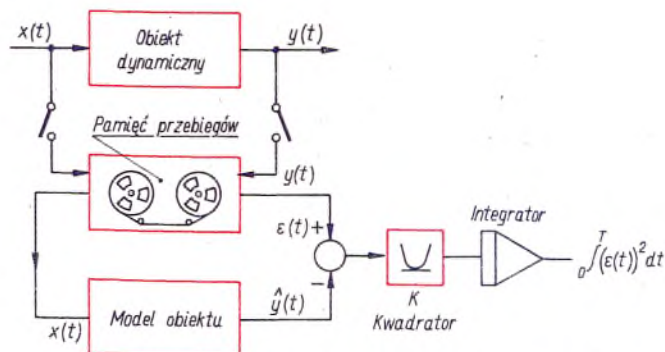
gdzie $\hat{y}(t)$ — sygnał wyjściowy modelu; $y(t)$ — sygnał wyjściowy obiektu.

Oczywiście wartość całki zależy od wartości parametrów modelu. Trzeba je więc tak dobrać, aby uzyskać jak najmniejszą wartość całki; model będzie wówczas najlepiej przybliżał (aproksymował) obiekt rzeczywisty (rys. 6.41B).

•6.41A• Identyfikacja parametryczna obiektu dynamicznego



•6.41B• Sposób wykonywania identyfikacji obiektu dynamicznego



Zmieniaj parametry modelu odtwarzając wielokrotnie sygnały $x(t)$, $y(t)$ tak, aby uzyskać minimum całki

Zadanie identyfikacji obiektu dynamicznego sprowadzone zostało do zadania optymalizacji

$$\min_a J(a)$$

gdzie a — współczynniki modelu.

•6.42• W jaki sposób dokonuje się identyfikacji „czarnej skrzynki”?

Jak wiadomo z teorii układów dynamicznych sygnał wyjściowy liniowego układu dynamicznego można przedstawić w postaci całki

$$y(t) = \int_0^t u(\tau)g(t-\tau)d\tau + y(0)$$

gdzie $u(t)$ jest sygnałem wejściowym; $g(t)$ — odpowiedzią impulsową obiektu.

W takiej właśnie postaci poszukiwać można modelu matematycznego „czarnej skrzynki”. Wielkością niewiadomą, którą należy określić na drodze identyfikacji jest funkcją $g(t)$.

Sposób postępowania jest analogiczny jak w przypadku identyfikacji parametrycznej układów dynamicznych. Rejestruje się sygnał wejściowy i wyjściowy obiektu rzeczywistego. Następnie zmierzony sygnał wejściowy podaje się na wejście modelu, otrzymując sygnał $\hat{y}(t)$. Całka określająca błąd modelu

$$J = \int_0^T [\hat{y}(t) - y(t)]^2 dt$$

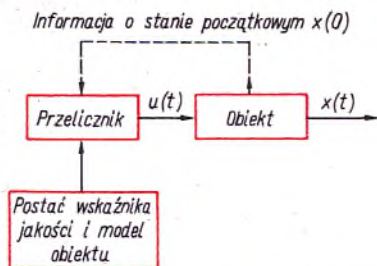
oczywiście zależy od funkcji $g(t)$. Można zatem sformułować następujące zadanie wariacyjne. Znaleźć funkcję $g(t)$ taką, aby uzyskać minimum wartości J . Oczywiście na funkcję $g(t)$ są nałożone pewne ograniczenia, powinien być np. spełniony warunek $g(t) = 0$ dla $t < 0$.

W wyniku rozwiązania tego zadania otrzymamy układ dynamiczny w najlepszy sposób przekształcający sygnały wejściowe na wyjściowe. Podkreślić należy fakt, że nie układaliśmy żadnych równań modelu — jedynym założeniem jest założenie o jego liniowości. Konsekwencją takiego założenia jest postać całki.

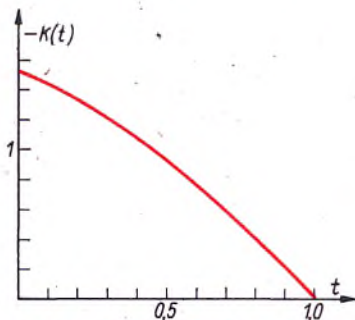
•6.43• Co to jest regulator optymalny?

Wynikiem rozwiązania rozważanych poprzednio zadań optymalizacji dynamicznej był optymalny przebieg sterowania $u(t)$. Rozwiązanie to zależało od warunków początkowych obiektu. Tak więc chcąc optymalnie sterować rzeczywisty obiekt należy:

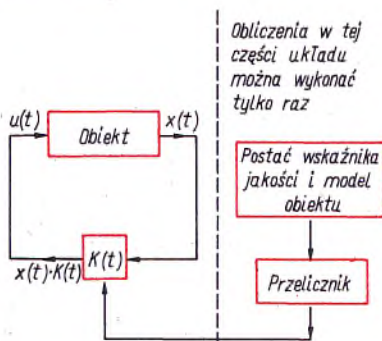
- Mierzyć warunki początkowe (stan) obiektu;
- Rozwiązywać zadanie optymalizacji dynamicznej;
- Obliczone sterowanie przyłożyć do obiektu.



•6.43A• Schemat optymalnego otwartego (programowego) układu sterowania



•6.43B• Wykres funkcji określającej wzmocnienie w pętli sprzężenia zwrotnego optymalnego zamkniętego układu sterowania



•6.43C• Schemat optymalnego zamkniętego układu sterowania

Tego typu układ optymalny (rys. 6.43A) nazywamy optymalnym otwartym lub programowym układem sterowania. Oczywiście, takie układy mogą być (i są) trudne w realizacji ze względu na dużą liczbę obliczeń niezbędnych do rozwiązania zadania optymalizacji. Celowe byłoby więc skonstruowanie optymalnego układu regulacji umożliwiającego wyznaczenie takiego rozwiązania, które można by stosować dla wszystkich możliwych warunków początkowych. Rozważmy dla ustalenia uwagi układ opisywany równaniem różniczkowym

$$\dot{x}(t) = u(t)$$

i wskaźnik jakości

$$J = \int_0^1 (x^2 + u^2) dt$$

Założmy, że warunek początkowy ma postać

$$x(0) = A$$

Można sprawdzić, że rozwiązanie optymalne jest następujące:

$$x(t) = A \frac{e^{1-t} + e^{-t}}{e^{-1} + e}, \quad u(t) = 2A \frac{-e^{1-t} + e^{-t}}{e^{-1} + e}$$

Sterowanie optymalne $u(t)$ zależy od warunków początkowych (stała A). Jednocześnie $x(t)$ też zależy od warunków początkowych. Dzieląc obie równości stronami otrzymamy

$$\frac{x(t)}{u(t)} = \frac{e^{1-t} + e^{-t}}{2(-e^{1-t} + e^{-t-1})}$$

Stąd wniosek: znając stan obiektu można wyznaczyć sterowanie optymalne z zależności

$$u(t) = K(t)x(t)$$

przy czym funkcja $K(t)$ **nie zależy od warunków początkowych** i można ją obliczyć tylko raz (rys. 6.43B). Z wniosku tego wynika recepta na realizację optymalnego układu sterowania: w każdej chwili czasu należy mierzyć współrzędne stanu obiektu, mnożyć ich wartość przez wartość funkcji $K(t)$ i otrzymamy sygnał przykładac na wejście obiektu (rys. 6.43C). Tego typu układy nazywamy **liniowymi regulatorami optymalnymi** (optymalnymi zamkniętymi układami sterowania).

•6.44• Czy zawsze można skonstruować regulator optymalny?

Okazuje się, że regulator optymalny można skonstruować zawsze, jeżeli obiekt sterowania jest liniowy i sterowalny, a wskaźnik jakości kwadratowy. Wówczas sterowanie optymalne zależy liniowo od stanu obiektu.

W bardziej złożonych przypadkach konstrukcja regulatora

może napotykać na bardzo znaczne trudności. W pewnych szczególnych przypadkach udaje się skonstruować takie regulatory, mają one jednak charakter układów nieliniowych.

•6.45• Jaki można podać przykład nieliniowego regulatora optymalnego?

Rozważmy walcownię-zgniatacz w hucie stali. Podczas walcowania porcja stali jest przepuszczana wielokrotnie między walcami aż do uzyskania odpowiednich rozmiarów. Przy każdym kolejnym przejściu między walcami odległość między nimi jest zmniejszana. Jeden cykl walcowania przebiega następująco:

- ustawienie walców
- uruchomienie układów napędowych „do przodu”
- zatrzymanie układów napędowych
- przestawienie walców
- uruchomienie układów napędowych „do tyłu”
- zatrzymanie układów napędowych

po czym cały cykl jest wielokrotnie powtarzany, aż do zakończenia procesu.

Zadaniem operatora walcowni jest wykonywanie opisanych wyżej czynności. Podczas wykonywania operacji przestawiania walców proces walcowania jest przerywany. Wynika stąd, że w celu uzyskania możliwie dużej wydajności walcowni operacja przestawiania walców powinna trwać jak najkrócej.

Zadanie przestawiania walców jest więc zadaniem sterowania optymalnego, w którym wskaźnikiem jakości jest czas trwania procesu. Tego typu zadania nazywamy **zadaniami czasowo-optymalnymi**.

Walce są przedstawiane za pomocą silnika elektrycznego. Jeżeli przez $u(t)$ oznaczmy wartość napięcia zasilającego silnik, a przez $x(t)$ — położenie regulowanego walca, to uproszczone równanie różniczkowe procesu będzie miało postać

$$\ddot{x}(t) = Ku(t)$$

gdzie K — pewna stała charakteryzująca właściwości procesu. Przyjmiemy, że operator ma do dyspozycji tylko trzy wartości sygnału sterującego $u(t)$:

1. u_{max} — silnik „cała naprzód”
2. $u_{min} = -u_{max}$ — silnik „cała wstecz”
3. 0 — silnik „stop”.

Równanie opisujące obiekt jest drugiego rzędu, zatem stan obiektu można w pełni scharakteryzować podając dwie wielkości — położenie walca względem pewnego punktu odniesienia (np. punktu docelowego), oraz prędkość przemieszczania walca względem tego punktu.

Założmy, że walec przemieszcza się z pewną prędkością. Włączenie silników „cała wstecz” spowoduje, że w pewnej chwili osiągnie on prędkość równą zeru. Jeżeli w tym samym



•6.46• Ilustracja zasady optymalności

Posługując się zasadą maksimum można wykazać, że sterowanie takie jest rzeczywiście optymalne.

•6.46• Co to jest zasada optymalności?

Założmy, że mamy do rozwiązania następujące zadanie: przejść z punktu A do punktu B minimalizując pewien wskaźnik jakości (funkcjonał) (rys. 6.46). Jest to zadanie wariacyjne (decyzją jest wybór odpowiedniej funkcji).

Założmy, że znaleźliśmy takie rozwiązanie. Zasada optymalności głosi:

Każdy końcowy odcinek drogi optymalnej jest również optymalny. Uzasadnienie tego faktu jest proste. Niech punkt A' leży na drodze optymalnej. Gdyby droga AB nie była optymalna, to istniałaby pewna inna droga łącząca te punkty i zapewniająca mniejszą wartość wskaźnika jakości. Ale wówczas chcąc przejść z punktu A do B opłacałoby się podzielić zadanie na dwa etapy — przejście z A do A' , i z A' do B po nowej drodze. Ogólne nakłady byłyby wówczas mniejsze. Jest to jednak niemożliwe, gdyż pierwotna droga AB była optymalna, a więc zapewniająca najmniejszą wartość wskaźnika jakości.

•6.47• Jaki można podać przykład zastosowania zasady optymalności?

Rozważmy zbiornik z pompą, która może tłoczyć lub pobierać wodę. Zadaniem jest napełnienie zbiornika od poziomu $h = 0$ (zbiornik pusty) do poziomu $h = 1$ (zbiornik pełny). Całkowity czas napełniania zbiornika wynosi 24 godziny, przy czym zmiany nastaw silnika pompy można dokonywać co 8 godzin (na początku każdej zmiany). W czasie całej zmiany nastawy pozostają niezmienione. Zakładamy, że koszt pracy pompy jest proporcjonalny do kwadratu przepływu, przy czym w czasie II zmiany energia elektryczna jest 2 razy tańsza, niż podczas zmiany I lub III (taryfa nocna). Należy znaleźć optymalne nastawy pompy.

Rozpatrzmy pracę systemu w trakcie III zmiany.

Koszt pracy pompy J_3 jest równy u_3^2 , gdzie u_3 — przepływ w ciągu III zmiany.

Mamy zatem

$$h_3 = h_2 + \alpha u_3$$

gdzie: h_2 — poziom wody w zbiorniku na początku III zmiany; α — współczynnik zależny od rozmiarów zbiornika. Tak więc

$$u_3 = \frac{h_3 - h_2}{\alpha}$$

Koszt pracy pompy w czasie III zmiany jest równy

$$J_3 = \left(\frac{h_3 - h_2}{\alpha} \right)^2$$

Analogicznie rozpatrujemy pracę pompy w ciągu II zmiany.

Otrzymujemy:

$$u_2 = \frac{h_2 - h_1}{\alpha} \quad J_2 = \frac{1}{2} \left(\frac{h_2 - h_1}{\alpha} \right)^2$$

Całkowity koszt pracy w czasie dwóch ostatnich zmian

$$J_{2,3} = J_2 + J_3 = \frac{1}{2} \left(\frac{h_2 - h_1}{\alpha} \right)^2 + \left(\frac{h_3 - h_2}{\alpha} \right)^2$$

Wielkość ta zależy od h_2 . Można zatem tak dobrać tę wartość, aby uzyskać minimalny koszt przejścia od zapełnienia do poziomu h_1 do zapełnienia do poziomu h_3 . Różniczkując otrzymamy

$$\hat{h}_2 = \frac{h_1 + 2h_3}{3}$$

oraz minimalny koszt

$$\hat{J}_{2,3} = \frac{1}{3} \left(\frac{h_3 - h_1}{\alpha} \right)^2$$

Jest to minimalny koszt, a więc zgodnie z zasadą optymalności trajektoria

$$h_1 \rightarrow \hat{h}_2 \rightarrow h_3$$

musi być fragmentem trajektorii optymalnej.

Rozpatrzmy działanie układu w czasie pierwszej zmiany. Całkowity koszt przejścia od poziomu h_1 do h_3

$$J_{1,3} = J_{1,2} + \hat{J}_{2,3}$$

Tak więc

$$J_{1,3} = \left(\frac{h_1 - h_0}{\alpha} \right)^2 + \frac{1}{3} \left(\frac{h_3 - h_1}{\alpha} \right)^2$$

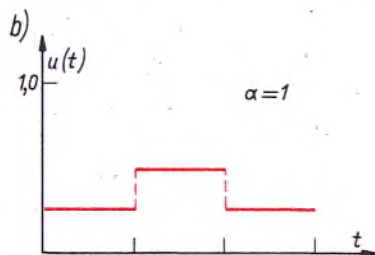
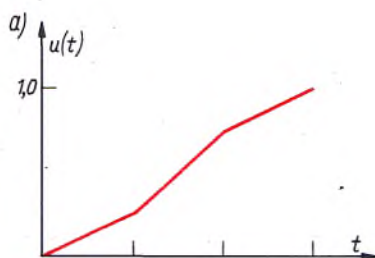
Możemy teraz dobrać h_1 tak, aby zminimalizować to wyrażenie. Po prostych rachunkach otrzymamy

$$\hat{h}_1 = \frac{3h_0 + h_3}{4}$$

Ponieważ w naszym przypadku $h_0 = 0$, $h_3 = 1$, to $h_1 = \frac{1}{4}$, $h_2 = \frac{3}{4}$, $h_3 = 1$. Stąd optymalne wartości przepływów (rys. 6.47)

$$u_1 = \frac{1}{4\alpha}, \quad u_2 = \frac{1}{2\alpha}, \quad u_3 = \frac{1}{4\alpha}$$

Zauważmy, że stosując zasadę optymalności rozwiązaliśmy zadanie niejako „od końca” — korzystaliśmy z faktu, że poszczególne odcinki trajektorii optymalnej też są optymalne. Umożliwiło to optymalizację poszczególnych etapów niezależnie — zadanie minimalizacji funkcji wielu zmiennych sprowadzono w ten sposób do szeregu kolejnych zadań minimalizacji funkcji jednej zmiennej.



•6.47• Optymalne napełnianie zbiornika: a) poziom wody w zbiorniku; b) natężenie dopływu wody

·7.1· Jakie maszyny matematyczne są obecnie stosowane?

Niewątpliwie najbardziej znanym typem maszyn matematycznych są maszyny cyfrowe. Poświęca się im wiele uwagi w literaturze popularno-naukowej, środkach masowego przekazu itp. Na skutek tego utożsamia się na ogół pojęcia „maszyna matematyczna” i „maszyna cyfrowa”. Tymczasem istnieją inne maszyny matematyczne o równie dużych, a nawet większych możliwościach. Co więcej, zakres zadań rozwiązywanych przez te maszyny i maszyny cyfrowe nie pokrywa się, a w wielu przypadkach okazują się one bardziej efektywne, niż maszyny cyfrowe.

Są to:

- maszyny analogowe
- maszyny iteracyjne
- maszyny hybrydowe.

Obecne tendencje rozwojowe wskazują na to, że w przyszłości w wielu dziedzinach zastosowań maszyny cyfrowe będą stopniowo zastępowane maszynami analogowymi, zarówno ze względu na ich mały koszt, jak i dużą efektywność działania.

·7.2· W jakich typowych problemach automatyki konieczne jest stosowanie maszyn matematycznych?

Zastosowanie maszyn matematycznych w automatyce można podzielić na trzy grupy:

Grupa pierwsza — to typowe obliczenia inżynierskie. Maszyna matematyczna używana jest w celu wykonania pewnej liczby obliczeń arytmetycznych, np. obliczanie wartości jakiegoś wzoru, lub tablicowanie funkcji. Tego typu zastosowaniami nie będziemy się zajmowali.

Druga grupa zadań, typowa dla automatyki, to zadania **symulacji**. Metody symulacyjne umożliwiają badanie układów automatyki na podstawie modeli matematycznych rozwiązywanych przez maszyny matematyczne. Możliwe jest np. symulowanie reaktora chemicznego wraz ze złożonym układem automatyki. Projektant układu regulacji może badać działanie projektowanego układu przy użyciu maszyny ma-

Tablica analogii

Układ mechaniczny

Masa m
Współczynnik oporu ρ
Stała sprężyny k
Siła ciężkości

Układ elektryczny

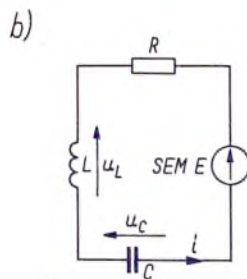
Indukcyjność L
Rezystancja R
Pojemność C
Siła elektromotoryczna

•7.3• Przykład analogii — układ elektryczny jest analogiem układu mechanicznego



Równanie stanu

$$m\ddot{x} + \rho\dot{x} + \frac{1}{k}x = mg$$



Równanie stanu

$$L \frac{d^2 i}{dt^2} + R \frac{di}{dt} + \frac{1}{C} i = \frac{dE}{dt}$$

tematycznej. Podkreślić należy, że może to być układ nie istniejący fizycznie — istnieje on tylko jako pewien program maszyny matematycznej.

Obecnie etap symulacji jest nieodzownym etapem projektowania układu regulacji. Umożliwia on szybkie i mało kosztowne zweryfikowanie prawidłowości wykonywanego projektu, doświadczalne dobranie parametrów (nastaw) regulatora itp.

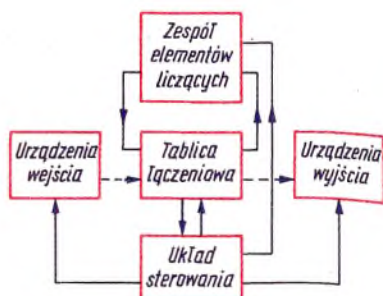
Trzecia grupa zastosowań — to maszyny matematyczne pełniące rolę urządzeń sterujących złożonymi systemami.

W niniejszym rozdziale zajmować się będziemy wyłącznie problemami symulacji układów dynamicznych.

•7.3• Co to są układy analogiczne (analogi)?

Jak wiadomo, istniejący fizycznie obiekt regulacji można opisać za pomocą układu równań różniczkowych. Równania te wiążą sygnały działające na obiekt z jego stanem, oraz sygnałami wyjściowymi. Jeżeli uda się znaleźć inny obiekt o **innej naturze fizycznej**, opisany identycznym układem równań, to obiekty te nazwiemy **analogicznymi** — jeden z nich może być **analogiem** (inaczej **modelem**) drugiego. Z faktu, że równania opisujące te obiekty są identyczne, wynika że obiekty te jednakowo reagują na takie same sygnały. Badanie obiektu można więc zastąpić badaniem jego analogu (modelu). Analogów należy więc poszukiwać w klasie takich systemów, które dają się w łatwy sposób zrealizować i badać. Najbardziej rozpowszechnione są analogi elektryczne (rys. 7.3). Są to układy elektryczne o odpowiednio dobranych parametrach, tak aby równania różniczkowe opisujące je miały pożądaną postać. Budowę analogów elektrycznych umożliwiają **analogowe maszyny matematyczne**.

•7.4• Jak jest zbudowana analogowa maszyna matematyczna?



•7.4• Schemat organizacji analogowej maszyny matematycznej

Analogowa maszyna matematyczna stanowi zbiór „cegiełek” umożliwiających budowę analogów. W zasadzie wszystkie istniejące maszyny analogowe są maszynami elektrycznymi.



•7.5• Symbol wzmacniacza operacyjnego

Maszyna składa się z pewnego zestawu elementów elektrycznych umożliwiających wykonanie elementarnych działań na sygnałach elektrycznych (dodawanie, mnożenie, całkowanie itp.). Podstawowym elementem umożliwiającym wykonywanie takich operacji jest **wzmacniacz operacyjny**. Wejścia i wyjścia tych elementów wyprowadzone są na płytę łączniową. Zadanie programisty budującego analog polega na odpowiednim połączeniu gniazdek za pomocą przewodów. Maszyna jest wyposażona ponadto w urządzenia ułatwiające pracę — układ sterowania elementami liczącymi oraz urządzenia umożliwiające odczyt i zapis wyników (rys. 7.4).

•7.5• Co to jest wzmacniacz operacyjny?

Wzmacniacz operacyjny jest to układ elektroniczny (obecnie wykonywany w postaci monolitycznego układu scalonego) wzmacniający napięcie. Wzmocnienie takiego wzmacniacza jest bardzo duże — rzędu 10^8 . Oznacza to, że sygnał (napięcie) na wyjściu wzmacniacza jest 10^8 razy większe niż na wejściu. Ponieważ wzmacniacz jest zasilany ze źródła o niskim napięciu (przeważnie 10 V), a napięcie wyjściowe nie może przekraczać wartości napięcia zasilania, to dopuszczalny sygnał wejściowy może zawierać się w granicach od -10^{-8} do 10^{-8} V. Oznacza to, że wzmacniacz powinien prawidłowo reagować na nadzwyczaj małe napięcia wejściowe. Powoduje to konieczność bardzo starannego zaprojektowania wzmacniacza. Wykonane są one tak, że nie pobierają praktycznie żadnego prądu wejściowego, rezystancja wyjściowa jest praktycznie równa zero (tzn. napięcie wyjściowe nie zależy od obciążenia wzmacniacza), a jego parametry bardzo nieznacznie zależą od napięcia zasilania i temperatury otoczenia.

Na rysunku 7.5 przedstawiono schematyczne oznaczenie wzmacniacza.

•7.6• Do czego służą wzmacniacze operacyjne?

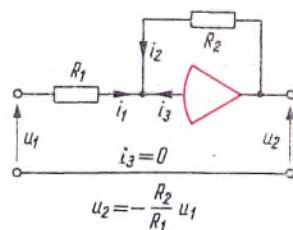
Przy użyciu wzmacniaczy operacyjnych można zbudować układy wykonujące pewne przekształcenia sygnałów elektrycznych. Przekształcenia te są następujące:

- Wzmacnianie sygnału napięciowego
- Sumowanie i odejmowanie napięć
- Przekształcenie sygnału napięciowego według zadanej zależności funkcyjnej
- Całkowanie sygnału napięciowego.

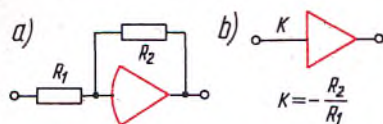
•7.7• W jaki sposób można zastosować wzmacniacz operacyjny do wzmacniania sygnału napięciowego?

Jak wiadomo, wzmocnienie wzmacniacza operacyjnego jest nadzwyczaj duże. W praktyce są potrzebne układy wzmacniające o wzmocnieniu znacznie mniejszym (rzędu od kilku do kilkudziesięciu). Do realizacji takiego układu można zastosować wzmacniacz operacyjny.

Rozważmy układ z rys. 7.7A. Jeżeli wzmacniacz nie pobiera



•7.7A• Wzmacniacz operacyjny użyty do wzmacniania sygnałów elektrycznych (inwerter)



•7.7B• Schemat inwertera i jego symbol

prądu (ma nieskończenie wielką rezystancję wejściową), to suma prądów dopływających do zacisku wejściowego wzmacniacza jest równa zero (II prawo Kirchhoffa). Podczas normalnej pracy wzmacniacza napięcie na jego wejściu jest nadzwyczaj małe. Możemy więc z dobrym przybliżeniem przyjąć, że jest ono równe zero. Tak więc

$$i_1 = \frac{u_1}{R_1}$$

$$i_2 = \frac{u_2}{R_2}$$

Ponieważ suma tych prądów ma być równa zero, to

$$i_1 + i_2 = \frac{u_1}{R_1} + \frac{u_2}{R_2} = 0$$

a zatem otrzymaliśmy

$$u_2 = -\frac{R_2}{R_1} u_1$$

Wniosek:

Wzmacniacz operacyjny o konfiguracji pokazanej na rys. 7.7B działa jak element proporcjonalny o współczynniku proporcjonalności $\frac{R_2}{R_1}$. Napięcie wyjściowe takiego układu ma przeciwny znak, niż napięcie wejściowe, a jego wartość chwilowa różni się $\frac{R_2}{R_1}$ razy.

•7.8• W jaki sposób można zastosować wzmacniacz operacyjny do sumowania napięć?

Rozważmy układ, którego schemat podano na rysunku 7.8. Jak poprzednio, zakładamy, że prąd wejściowy wzmacniacza jest równy zero, a napięcie wejściowe pomijalnie małe. Tak więc suma prądów dopływających do wejścia wzmacniacza jest równa zero (II prawo Kirchhoffa). Prądy te są równe

$$i_1 = \frac{u_1}{R_1}, \quad i_2 = \frac{u_2}{R_2}, \quad i_3 = \frac{u_3}{R_3}$$

sumując je otrzymamy równanie

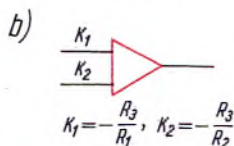
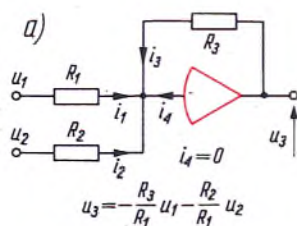
$$i_1 + i_2 + i_3 = \frac{u_1}{R_1} + \frac{u_2}{R_2} + \frac{u_3}{R_3} = 0$$

Możemy więc obliczyć napięcie wyjściowe wzmacniacza

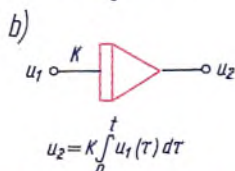
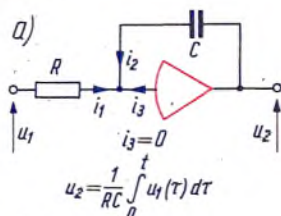
$$u_3 = -u_1 \frac{R_3}{R_1} - u_2 \frac{R_3}{R_2}$$

Wniosek:

Wzmacniacz operacyjny pracujący w układzie pokazanym na rys. 7.8 wykonuje następujące operacje:



•7.8• Zastosowanie wzmacniacza operacyjnego do sumowania sygnałów elektrycznych: a) realizacja; b) symbol



·7.9· Wzmacniacz operacyjny użyty do całkowania sygnałów elektrycznych (integrator): a) realizacja; b) symbol

— Mnoży napięcie wejściowe przez współczynniki zależne od wartości rezystancji

— Sumuje te napięcia

— Zmienia znak wyniku.

Wzmacniacz oblicza więc **sumę ważoną napięć wejściowych**. W szczególnym przypadku, gdy rezystancje R_1, R_2, R_3 są równe, wzmacniacz po prostu sumuje napięcia wejściowe.

·7.9· W jaki sposób można zastosować wzmacniacz operacyjny do całkowania napięcia zmiennego w czasie?

Zastosowanie wzmacniacza operacyjnego do całkowania napięcia zmiennego w czasie pokazano na rys. 7.9. W układzie tym wejście jest połączone z wyjściem wzmacniacza za pośrednictwem kondensatora o pojemności C .

Przyjmujemy, jak poprzednio, że prąd wejściowy i napięcie wejściowe wzmacniacza są pomijalnie małe. Z elektrotechniki wiadomo, że

$$i_c(t) = C \frac{du_2(t)}{dt}$$

Korzystając z II prawa Kirchhoffa otrzymamy

$$i_1(t) + i_c(t) = \frac{u_1(t)}{R} + C \frac{du_2(t)}{dt} = 0$$

Stąd

$$\frac{du_2(t)}{dt} = -\frac{1}{RC} u_1(t)$$

Całkując obustronnie powyższą równość otrzymamy

$$u_2(t) = -\frac{1}{RC} \int_0^t u_1(\tau) d\tau + u_2(0)$$

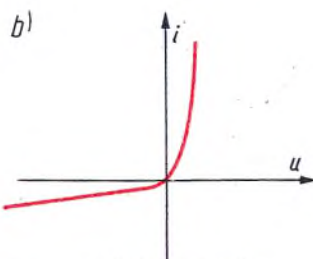
Z zależności powyższej wynika, że napięcie na wyjściu wzmacniacza jest proporcjonalne do całki napięcia wejściowego. Z tego względu układ powyższy nazywa się układem całkującym (integratorem). Schematyczne oznaczenie układu podano na rys. 7.9b.

·7.10· Jakie najprostsze funkcje nieliniowe można zrealizować za pomocą wzmacniacza operacyjnego?

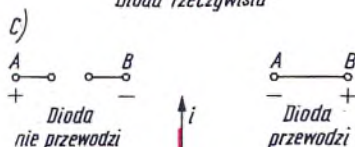
W wielu przypadkach zachodzi konieczność budowy tzw. przekształtników funkcyjnych realizujących zależność

$$u_2 = f(u_1)$$

Najprostsze zależności funkcyjne można zrealizować przy użyciu **diody** (rys. 7.10A). Diodę można traktować jako rezystor, którego rezystancja jest nieskończenie duża, jeżeli napięcie dodatnie jest na końcówce A, a ujemne na B, oraz którego rezystancja jest zerowa w przeciwnym przypadku. Oczywiście, dioda rzeczywista jest tylko pewnym przybliżeniem opisanej diody (idealnej). W praktyce w wielu przy-

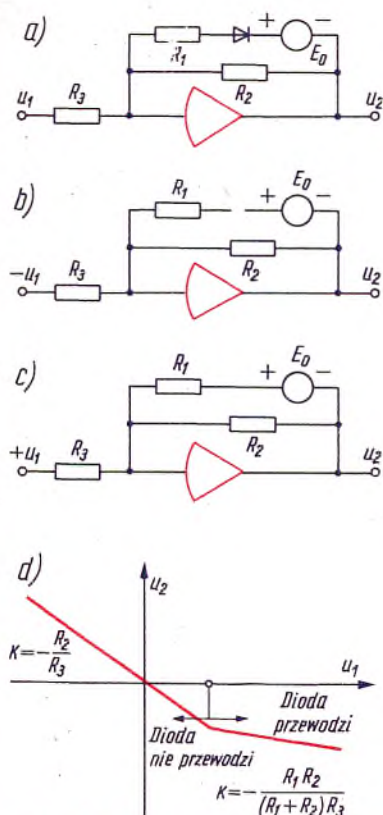


Dioda rzeczywista



Dioda idealna

·7.10A· Symbol i charakterystyki diody rzeczywistej i idealnej



•7.10B• Realizacja charakterystyki odcinkami liniowej za pomocą wzmacniacza operacyjnego i diody w pętli sprzężenia zwrotnego: a) schemat układu; b) dioda nie przewodzi; c) dioda przewodzi; d) charakterystyka układu

padkach traktować możemy diodę jako idealną. Rozpatrzmy dwa rodzaje układów nieliniowych — z diodą w pętli sprzężenia zwrotnego, i z diodą w obwodzie wejściowym. W obu układach dioda połączona jest szeregowo z rezystorem i źródłem napięciowym włączonym tak, aby nie przewodziła przy zerowym napięciu zewnętrznym (rys. 7.10B(a)). Załóżmy, że na wejście wzmacniacza podajemy napięcie ujemne. Dioda wówczas nie przewodzi, jej rezystancja jest nieskończenie duża. Wejście wzmacniacza jest połączone z wyjściem za pośrednictwem rezystora R_2 (rys. 7.10B(b)). Układ pracuje jak wzmacniacz o wzmocnieniu

$$K = -\frac{R_2}{R_3}$$

Następnie zmieniamy znak napięcia wejściowego. Jeżeli napięcie wejściowe jest niewielkie, to dioda również jest spolaryzowana w kierunku zaporowym. Efektywne napięcie działające na diodę

$$E_D = -\frac{R_2}{R_3} u_1 + E_0$$

gdzie E_0 — napięcie dodatkowego źródła. Gdy napięcie E_D zmienia znak (tzn. $E_D < 0$), dioda zaczyna przewodzić. Przy napięciu wejściowym większym od

$$u_1 = E_0 \frac{R_3}{R_2}$$

wejście i wyjście wzmacniacza są połączone za pomocą dwóch rezystorów R_1 i R_2 (rys. 7.10B(c)). Wypadkowa ich rezystancja

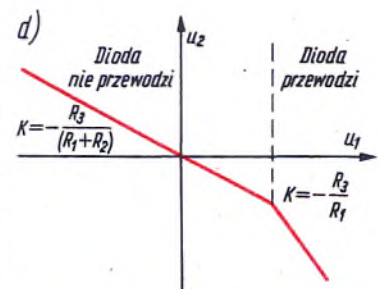
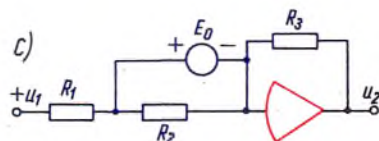
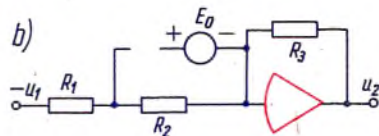
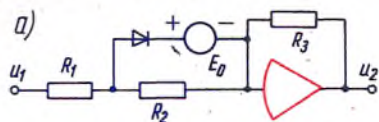
$$R_4 = \frac{R_1 R_2}{R_1 + R_2}$$

a wzmocnienie całego układu

$$K = -\frac{R_1 R_2}{(R_1 + R_2) R_3}$$

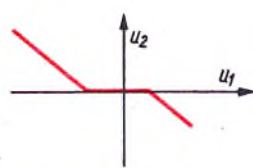
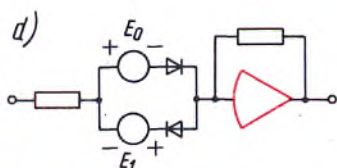
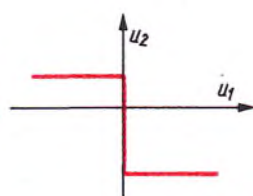
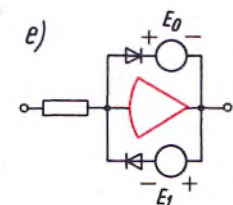
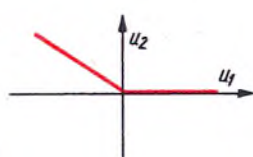
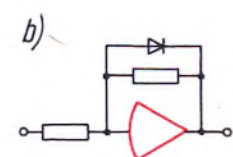
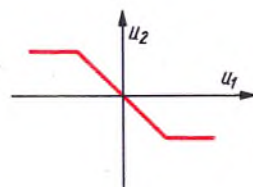
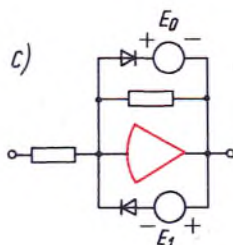
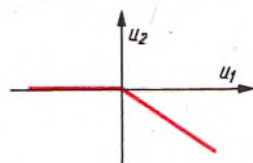
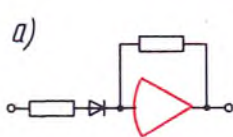
Rozważany układ realizuje charakterystykę **odcinkami liniową** (łamaną) (rys. 7.10B(d)). Nachylenie obu odcinków łamanej można zmieniać dobierając odpowiednio rezystancje R_1 i R_2 , natomiast zmiana napięcia E_0 powoduje przesuwanie punktu załamania.

Analiza układu z diodą w obwodzie wejściowym (rys. 7.10C) jest bardzo podobna i nie będziemy jej tu powtarzać. Na rysunku 7.10D przedstawiono najczęściej spotykane układy nieliniowe, które można zrealizować przy użyciu wzmacniaczy operacyjnych i diod. Analizę działania tych układów pozostawiamy Czytelnikowi.



•7.10C• Realizacja charakterystyki odcinkami liniowej wzmacniacza operacyjnego i diody w obwodzie wejściowym wzmacniacza za pomocą: a) schemat układu; b) dioda nie przewodzi; c) dioda przewodzi; d) charakterystyka układu

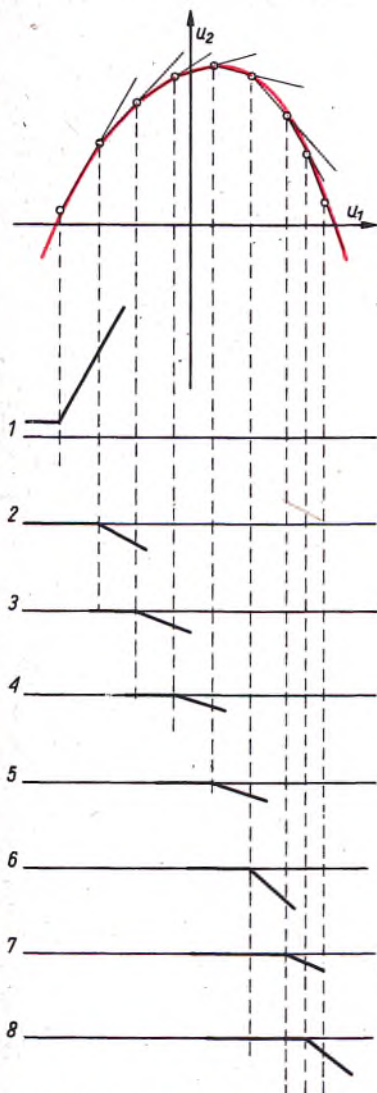
•7.10D• Charakterystyki możliwe do uzyskania za pomocą układów diodowych



•7.11• W jaki sposób realizuje się dowolne zależności funkcyjne (np. $y = x^2$)?

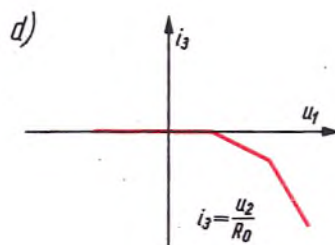
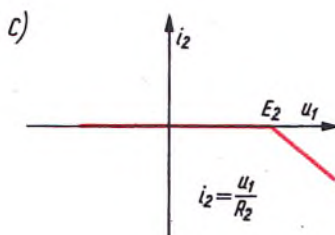
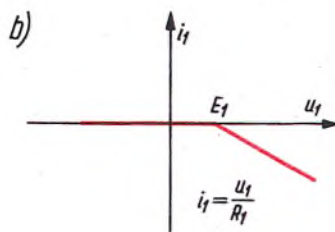
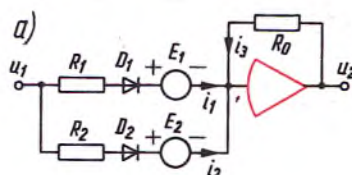
Bardzo często zachodzi potrzeba realizacji dowolnych zależności funkcyjnych zadanych analitycznie lub w postaci wykresu otrzymanego w wyniku pomiarów. Funkcje takie można realizować korzystając z omówionych poprzednio układów zawierających diody. Sposób postępowania jest następujący (rys. 7.11A):

- Sporządza się dokładny wykres funkcji;
- Przedział zmienności zmiennej niezależnej dzieli się równo odległymi punktami na pewną liczbę (na ogół 10 ... 20) odcinków;
- Oblicza się wartości funkcji w tych punktach;
- Dokładny wykres funkcji przybliża się funkcją łamaną (funkcją odcinkami liniową).



•7.11A• Sposób przybliżania funkcji ciągłej funkcją łamaną i rozkład funkcji łamanej na elementarne składowe

•7.11B• Przekształtnik funkcyjny: a) uproszczony schemat układu; b, c) — zależność prądu przepływającego przez elementarne układy wejściowe od napięcia wejściowego; d) prąd sumaryczny przepływający w kierunku wejścia wzmacniacza



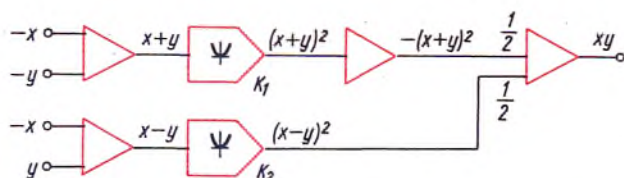
Można teraz przedstawić funkcję w postaci sumy funkcji łamanych z jednym punktem załamania, łatwych do realizacji. Na rysunku 7.11B przedstawiono przykład realizacji takiego układu. Oczywiście, układy stosowane w praktyce są bardziej złożone, jednak zasada ich działania jest taka sama.

•7.12• W jaki sposób można mnożyć sygnały elektryczne?

Mnożenie sygnałów jest jedną z najczęściej spotykanych operacji nieliniowych. Istnieje wiele metod realizacji układu mnożącego; omówimy najprostszy z nich. Łatwo sprawdzić następującą tożsamość

$$xy = \frac{1}{2} [(x+y)^2 - (x-y)^2]$$

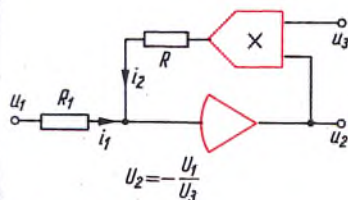
•7.12• Uproszczony schemat układu mnożącego



Tak więc do realizacji mnożenia należy dysponować pewną liczbą sumatorów, oraz dwoma układami realizującymi podnoszenie do kwadratu (tzw. kwadratory). Układy te realizuje się na ogół stosując przybliżenie funkcji $y = x^2$ funkcją odcinkami liniową (łamaną). Realizację takich układów omówiono w odpowiedzi na poprzednie pytanie. Dokładny schemat układu realizującego powyższą zależność przedstawiono na rys. 7.12. Układy stosowane w praktyce są znacznie bardziej złożone, lecz wymagają tylko jednego wzmacniacza operacyjnego. Ich zasada działania jest jednak taka sama, jak omówionego układu.

•7.13• W jaki sposób można dzielić sygnały elektryczne?

Dysponując członem mnożącym można w łatwy sposób dzielić sygnały. Wystarczy w tym celu połączyć wejście i wyjście wzmacniacza za pośrednictwem członu mnożącego (rys. 7.13). Jeżeli wzmacniacz ma nieskończenie dużą rezystancję wejściową, a jego napięcie wejściowe jest pomijalnie małe, to możemy napisać



•7.13• Układ dzielnicy

$$i_1 + i_2 = \frac{u_1}{R} + \frac{u_2 u_3}{R} = 0$$

Wynika stąd, że

$$u_2 = -\frac{u_1}{u_3}$$

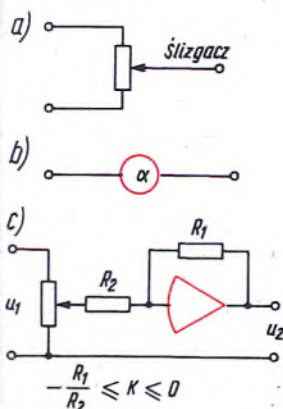
a więc układ realizuje operację dzielenia dwóch sygnałów elektrycznych.

•7.14• Co to jest potencjometr i do czego służy?

Bez tego elementu trudno byłoby zrealizować na maszynie analogowej jakikolwiek program. Jest to rezystor z odczepem, którego położenie można dowolnie zmieniać (rys. 7.14). Przyłączając do końcówek potencjometru źródło napięciowe o napięciu E można, przesuwając ślizgacz (odczep) potencjometru, ustawiać na nim dowolne napięcie zawierające się w przedziale $[0, E]$. Przyłączając potencjometr do wejścia wzmacniacza operacyjnego pracującego jako wzmacniacz można regulować wartość wzmocnienia (rys. 7.14b). W analogiczny sposób można zmieniać współczynnik całkowania integratora.

Konieczność stosowania potencjometrów wynika z faktu, że wszystkie rezystory stosowane w maszynie analogowej mają standardowe wartości rezystancji, np. 1, 10, 100 kΩ. Tak

•7.14• Potencjometr: a) schemat; b) symbol; c) zastosowanie potencjometru do zmiany wzmocnienia inwertera



więc dysponując tymi rezystorami można zrealizować tylko niewielką liczbę kombinacji (z przyczyn technicznych nie ma na ogół możliwości łączenia rezystorów szeregowo lub równolegle). Aby otrzymać wartość wzmacnienia różną od wartości wynikających z kombinacji standardowych elementów należy zastosować potencjometr.

•7.15• Co to jest program maszyny analogowej?

Maszyna analogowa służy do symulacji układów dynamicznych opisywanych równaniami różniczkowymi zwyczajnymi. Załóżmy więc, że mamy sformułowany problem: rozwiązać równanie różniczkowe, z zadanymi warunkami początkowymi.

Program maszyny analogowej jest to sposób połączenia bloków maszyny tak, aby otrzymany w wyniku połączenia układ elektryczny był opisany takim samym równaniem różniczkowym jak to, które mamy rozwiązać.

Programowanie maszyny analogowej składa się z trzech etapów:

- Ustalenie schematu połączeń;
- Ustalenie wartości elementów (dobór rezystancji, pojemności, napięć, nastaw potencjometrów);
- Połączenie schematu.

•7.16• Jakie najprostsze równanie różniczkowe można rozwiązać za pomocą maszyny analogowej?

Omawiając działanie integratora stwierdziliśmy, że napięcia wejściowe i wyjściowe są związane następującą zależnością

$$\frac{du_2}{dt} = -\frac{1}{RC} u_1$$

Tak więc, jest to najprostsze równanie różniczkowe jakie można rozwiązać za pomocą maszyny analogowej (rys. 7.16).

•7.17• W jaki sposób rozwiązuje się złożone równania różniczkowe pierwszego rzędu?

Załóżmy, że mamy rozwiązać równanie

$$b \frac{dx(t)}{dt} + cx(t) = u(t)$$

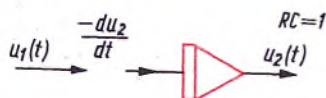
Przekształcając to równanie otrzymamy

$$\frac{dx}{dt} = -\frac{c}{b} x(t) + \frac{u(t)}{b}$$

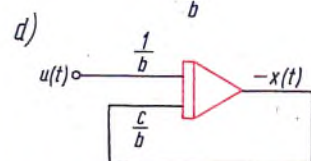
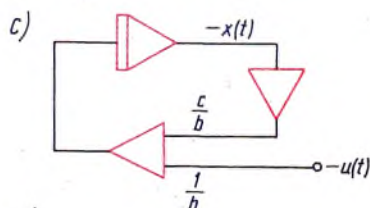
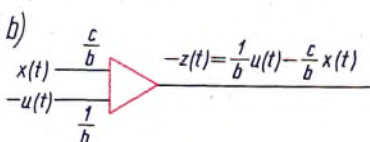
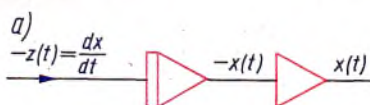
Jeżeli oznaczmy

$$-\frac{c}{b} x(t) + \frac{u(t)}{b} = -z(t)$$

to równanie możemy zapisać w następującej postaci



•7.16• Integrator jako układ rozwiązujący najprostsze równanie różniczkowe



•7.17• Etapy programowania równania różniczkowego pierwszego rzędu ($b, c > 0$). Schemat (c) można uprościć korzystając z faktu, że dwukrotne przejście sygnału przez wzmacniacz operacyjny nie zmienia znaku sygnału

$$\frac{dx}{dt} = -z(t)$$

Wiemy, że równanie takie można rozwiązać za pomocą tylko jednego integratora. Jeżeli na wejście tego integratora podamy sygnał $-z(t)$, a stałe R i C dobierzemy tak, aby $RC = 1$, to napięcie wyjściowe integratora będzie spełniało równanie, a więc będzie równe $x(t)$. Pozostaje problem określenia funkcji $z(t)$ (na razie nieznanej). Funkcję tę łatwo wyznaczyć, znając $x(t)$ i $u(t)$. Kolejne etapy formułowania programu podano na rys. 7.17.

•7.18• W jaki sposób rozwiązuje się najprostsze równanie różniczkowe drugiego rzędu?

Najprostsze równanie różniczkowe drugiego rzędu ma postać

$$\frac{d^2x(t)}{dt^2} = u(t)$$

Równanie to można zapisać następująco:

$$\frac{d}{dt} \left(\frac{dx(t)}{dt} \right) = u(t)$$

Oznaczając

$$\frac{dx(t)}{dt} = z(t)$$

otrzymamy znane nam równanie różniczkowe (pyt. 7.17)

$$\frac{dz(t)}{dt} = u(t)$$

Tak więc zastąpiliśmy równanie różniczkowe drugiego rzędu układem dwóch równań różniczkowych pierwszego rzędu. Każde z tych równań bez trudności daje się rozwiązać za pomocą jednego integratora. Kolejne etapy formułowania programu pokazuje rys. 7.18.

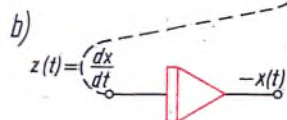
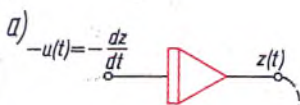
•7.19• W jaki sposób rozwiązuje się równania wyższego rzędu?

Zaprogramujemy równanie różniczkowe drugiego rzędu o postaci

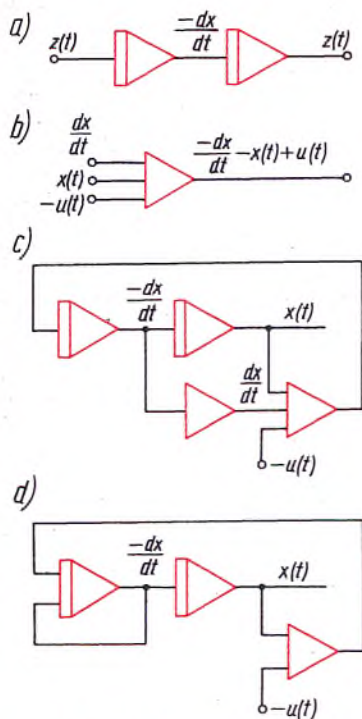
$$\frac{d^2x(t)}{dt^2} + \frac{dx(t)}{dt} + x(t) = u(t) \quad (7.19.1)$$

Zrozumienie tego przykładu pozwoli Czytelnikowi programować bardziej złożone równania. Przekształćmy równanie do postaci

$$\frac{d^2x(t)}{dt^2} = -\frac{dx(t)}{dt} - x(t) + u(t)$$



•7.18• Rozwiązanie najprostszego równania różniczkowego drugiego rzędu za pomocą dwóch integratorów



·7.19· Etapy programowania równania różniczkowego drugiego rzędu. Schemat (c) można uprościć korzystając z faktu, że dwukrotne przejście sygnału przez wzmacniacz operacyjny nie zmienia znaku sygnału

Jeżeli oznaczmy

$$z(t) = -\frac{dx(t)}{dt} - x(t) + u(t)$$

to równanie (7.19.1) możemy zapisać

$$\frac{d^2x(t)}{dt^2} = z(t)$$

Wiemy już, że równanie to można rozwiązać za pomocą dwóch integratorów. Sygnały wyjściowe integratorów są następujące:

$$\text{Integrator I} \quad -\frac{dx(t)}{dt}$$

$$\text{Integrator II} \quad x(t)$$

Pozostaje zatem określić sygnał $z(t)$. Można to zrobić w łatwy sposób na podstawie znajomości sygnałów $x(t)$ oraz $\frac{dx(t)}{dt}$. Kolejne etapy formułowania programu pokazuje rys. 7.19.

·7.20· Co to jest skalowanie maszyny analogowej?

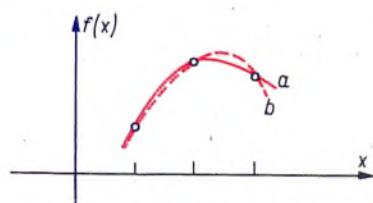
Skalowanie jest jednym z etapów programowania maszyny analogowej. Modelowane równania mogą opisywać zachowanie się bardzo różnych układów fizycznych. Załóżmy dla ułatwienia uwagi, że modelujemy reaktor chemiczny. Stan takiego reaktora można scharakteryzować podając dwie wielkości: stężenie reagenta i temperaturę. Wartość stężenia może zmieniać się od 0 do 1, natomiast temperatura — od 20 do 500°C.

Tymczasem w maszynie analogowej wszystkie zmienne reprezentowane są przez napięcia, które nie mogą przekraczać pewnych granicznych wartości (na ogół ± 10 V). Powstaje więc konieczność ustalenia pewnych współczynników proporcjonalności wiążących zmienne maszyny (napięcia) ze zmiennymi fizycznymi (np. temperatura). Czynność tę nazywamy skalowaniem. Współczynniki te muszą być ponadto tak dobrane, aby podczas pracy maszyny (symulacji) żadne z napięć nie osiągnęło ani przez chwilę wartości granicznej (+ lub - 10 V). W przeciwnym przypadku otrzymamy rozwiązanie obciążone bardzo dużym błędem.

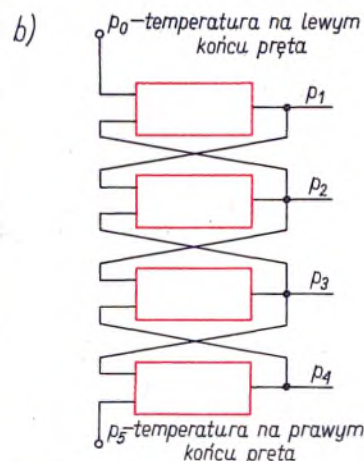
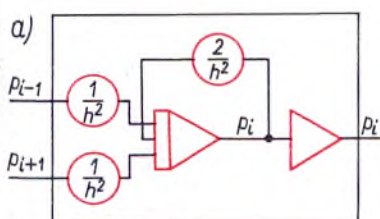
Wyznaczenie współczynników wymaga więc znajomości zakresu, w jakim zmieniają się wartości zmiennych fizycznych. Ponieważ zakres ten może być początkowo nieznany, skalowanie przeprowadza się wielokrotnie metodą prób i błędów.

·7.21· Czy za pomocą maszyny analogowej można modelować układy opisywane równaniami różniczkowymi cząstkowymi?

Można. Podobnie jak w przypadku zadanie optymalizacji należy problem rozwiązywania równania cząstkowego spro-



·7.21A· Przybliżenie funkcji funkcją kwadratową a) funkcja; b) parabola przybliżająca funkcję



·7.21B· Modelowanie analogowe układu opisanego równaniem różniczkowym cząstkowym (przepływ ciepła w pręcie): a) schemat jednego ogniwa modelu; b) sposób łączenia ogniw

wadzić do problemu rozwiązywania równania różniczkowego zwyczajnego. Zrobić to można stosując **metodę dyskretyzacji**.

Rozważmy równanie opisujące rozkład temperatury w pręcie

$$\frac{\partial p}{\partial t} = \frac{\partial^2 p}{\partial x^2}$$

Funkcja p zależy zarówno od czasu t , jak i od zmiennej przestrzennej x .

Rezygnujemy ze znajomości wartości funkcji w każdym punkcie x , zadowolimy nas znajomości jej wartości w pewnych wybranych punktach $x_1, x_2 \dots$ itd. Załóżmy, że punkty te są równo odległe. Możemy teraz wyznaczyć pochodną funkcji $p(x, t)$ w punkcie x_i , na podstawie znajomości jej wartości w trzech punktach — x_{i-1}, x_i, x_{i+1} .

Przez trzy punkty można zawsze poprowadzić parabolę o równaniu

$$p(x) = ax^2 + bx + c$$

Parabola ta spełnia warunki:

$$p(x_{i-1}) = ax_{i-1}^2 + bx_{i-1} + c$$

$$p(x_i) = ax_i^2 + bx_i + c$$

(7.23.1)

$$p(x_{i+1}) = ax_{i+1}^2 + bx_{i+1} + c$$

w których $p(x_{i-1}), p(x_i), p(x_{i+1})$ — wartości funkcji będącej rozwiązaniem równania w chwili t . Jeśli punkty są niezbyt odległe, to parabola przybliża funkcję $p(x, t)$ (rys. 7.21A).

Można zatem napisać

$$\frac{d^2 p}{dx^2} = 2a, \quad \text{w punkcie } x = x_i$$

Chcąc więc wyznaczyć drugą pochodną funkcji na podstawie znajomości jej wartości w trzech sąsiednich punktach należy rozwiązać układ równań (7.23.1). W wyniku prostych rachunków otrzymujemy

$$\frac{d^2 p}{dx^2} = \frac{p(x_{i-1}) - 2p(x_i) + p(x_{i+1}))}{h^2}$$

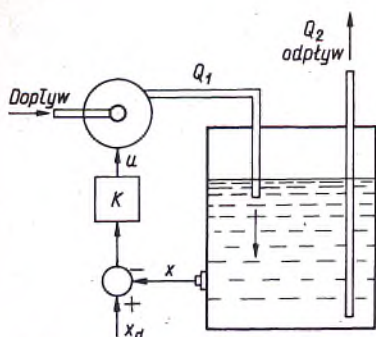
Przy czym $h = x_i - x_{i-1}$ — odległość między dwoma kolejnymi punktami.

Możemy więc napisać układ równań różniczkowych zwyczajnych

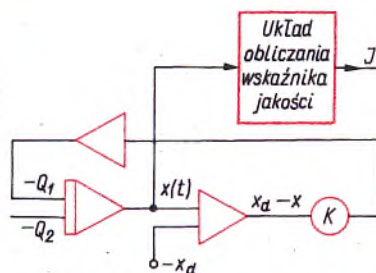
$$\frac{dp(x_i, t)}{dt} = \frac{p(x_{i-1}, t) - 2p(x_i, t) + p(x_{i+1}, t))}{h^2}$$

w przybliżeniu równoważny równaniu cząstkowemu.

Równania takie należy napisać dla wszystkich punktów, w których poszukujemy wartości funkcji p . Można je



·7.22A· Schemat układu regulacji poziomu



·7.22B· Model analogowy układu regulacji poziomu



·7.22C· Schemat blokowy procesu doboru optymalnych nastaw (optymalizacji parametrycznej) regulatora

bez kłopotu modelować za pomocą maszyny analogowej. Na rysunku 7.21B przedstawiono schemat (program) analogowy do wyznaczania rozkładu temperatury w przecie (a właściwie wartości tej temperatury w 4 wybranych punktach), którego końce są utrzymywane w stałej temperaturze.

·7.22· Czy można zastosować maszynę analogową do syntezy układu regulacji?

Analizę możliwości zastosowań maszyny analogowej przedstawimy na prostym przykładzie.

Założmy, że mamy dobrać nastawy (wzmocnienie) regulatora poziomu cieczy w zbiorniku (rys. 7.22A). Regulator ten steruje obroty pompy doprowadzającej lub odprowadzającej wodę ze zbiornika. Zadanie doboru nastaw regulatora można sformułować jako zadanie optymalizacji — nastawy trzeba dobrać tak, aby uzyskać minimum pewnego wskaźnika jakości. Można w prosty sposób zbudować analogowy model układu regulacji. Nieco bardziej złożony jest układ obliczający wskaźnik jakości; nie będziemy go szczegółowo rozważać. Zauważmy, że współczynnik wzmocnienia regulatora jest w naszym układzie reprezentowany przez **współczynnik potencjometru** (rys. 7.22B). Wykonując symulację wielokrotnie, przy różnych nastawach potencjometru można znaleźć nastawy regulatora zapewniające minimum wskaźnika jakości.

Na rysunku 7.22C przedstawiono w postaci schematu blokowego kolejność czynności niezbędnych do otrzymania rozwiązania zadania. Jak widać, tylko jedną z tych czynności można wykonać za pomocą maszyny analogowej, pozostałe — wybór nowej wartości wzmocnienia i podjęcie decyzji o zakończeniu obliczeń musi wykonać operator. Fakt ten powoduje, że proces wyznaczania rozwiązania może trwać bardzo długo (mimo dużej szybkości pracy maszyny analogowej).

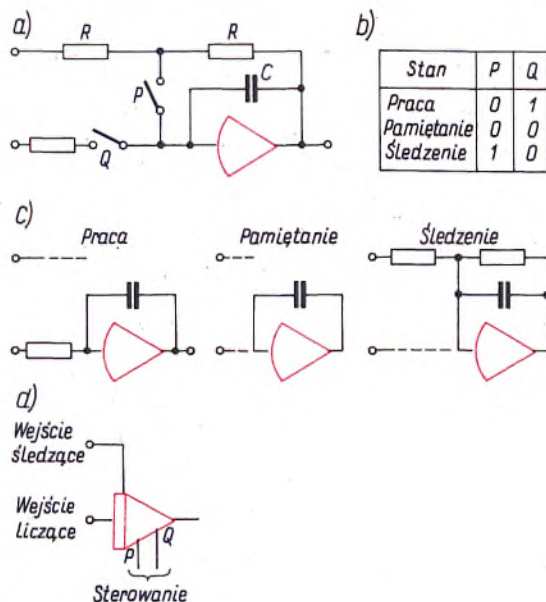
Zautomatyzowanie wszystkich czynności operatora zapewniają maszyny iteracyjne.

·7.23· Co to jest integrator sterowany?

Rozważany poprzednio integrator (pyt. 7.9) nie nadaje się do bezpośredniego zastosowania. Istotnie, natychmiast po włączeniu zasilania rozpocząłby pracę, niemożliwe byłoby zatrzymanie całkowania (np. w celu dokonania pomiarów) lub wprowadzenie warunków początkowych. W celu umożliwienia realizacji tych czynności w maszynie analogowej stosuje się **integratory sterowane**.

Na rysunku 7.23 przedstawiono schemat integratora sterowanego oraz zasadę jego działania. Integrator może pozostawać w jednym z trzech stanów:

- stan „praca”
- stan „pamiętanie”
- stan „wprowadzenie warunków początkowych” („śledzenie”).



·7.23· Integrator sterowany; a) schemat połączeń; b) tabelka stanów; c) schematy zastępcze układu w poszczególnych stanach; d) symbol

Każdy ze stanów można uzyskać ustawiając odpowiednio dwa przełączniki.

W stanie **praca** integrator całkuje sygnały podawane na jedno z jego wejść (wejście liczące).

W stanie **pamiętanie** wszystkie wejścia są odłączane od integratora — napięcie wyjściowe równe jest napięciu na kondensatorze integratora. Ponieważ wzmacniacz operacyjny nie obciąża kondensatora (prąd wejściowy jest równy zero), to napięcie to może się utrzymywać bez zmian przez bardzo długi okres czasu.

W stanie **wprowadzanie warunków początkowych** (śledzenie) wejście liczące jest odłączone, jest natomiast przyłączone wejście **śledzące**. Integrator działa teraz jak układ rozwiązujący równanie różniczkowe

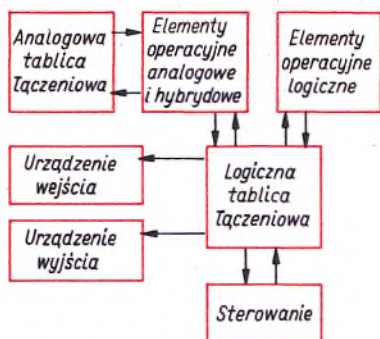
$$RC\dot{x} + x = u$$

Gdy sygnał u jest stały, to po dostatecznie długim czasie układ osiągnie stan ustalony, pochodna $\dot{x}$ będzie równa zero i z równania otrzymamy $x = u$. Tak więc napięcie wyjściowe integratora w stanie ustalonym jest równe napięciu na wejściu śledzącym. Wartości rezystora R dobiera się tak, aby czas ustalania napięcia wyjściowego był bardzo krótki.

W maszynie analogowej wszystkie integratory sterowane są jednocześnie, tzn. wszystkie znajdują się w tym samym stanie; wyboru stanu dokonuje się wciskając odpowiednie przyciski na tablicy sterującej.

·7.24· Co to jest maszyna iteracyjna?

Maszyna iteracyjna jest to maszyna analogowa, w której wszystkie integratory mogą być sterowane niezależnie.



·7.24· Schemat organizacji iteracyjnej maszyny matematycznej

W maszynie konwencjonalnej sterowanie integratorami odbywało się za pośrednictwem centralnego układu sterowania. Wciśnięcie klawisza, np. „warunki początkowe” powodowało przejście w s z y s t k i c h integratorów do tego stanu. W maszynie iteracyjnej istnieje możliwość sterowania indywidualnego. Dodatkowo maszyna taka jest wyposażona w układy logiczne umożliwiające **programowanie sterowania**, czyli wytworzenie odpowiednich impulsów sterujących grupy integratorów lub poszczególne integratory. Schemat organizacji maszyny iteracyjnej pokazano na rys. 7.24.

·7.25· Co to jest program maszyny iteracyjnej?

Na pytanie to odpowiemy korzystając z przykładu zadania doboru nastaw regulatora. Sprowadza się ono do następujących czynności:

- Wybór współczynnika wzmocnienia regulatora;
- Ustawienie potencjometru;
- Rozwiązanie równania opisującego układ regulacji (symulacja);
- Zapamiętanie wyniku;
- Podjęcie decyzji, czy uzyskane rozwiązanie jest zadowalające.

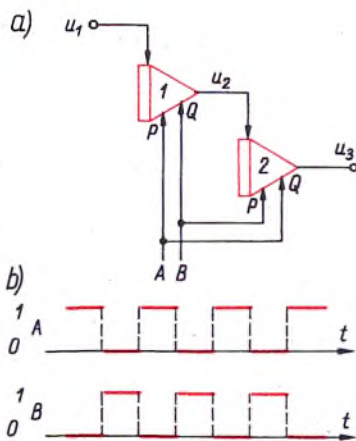
Czynności te stanowią **sieć działań** (czynności) maszyny iteracyjnej. Ustalenie sieci jest pierwszym etapem programowania. Następne etapy są następujące:

- Ustalenie programu części analogowej (część symulacyjna programu);
- Ustalenie sposobu wyboru współczynnika wzmocnienia (część arytmetyczna programu);
- Ustalenie sposobu pracy układów pamięciowych;
- Ustalenie sposobu zatrzymania programu;
- Opracowanie kolejności pracy integratorów;
- Opracowanie układu logicznego generującego impulsy sterujące dla integratorów;
- Połączenie programu na logicznej i analogowej tablicy łączeniowej.

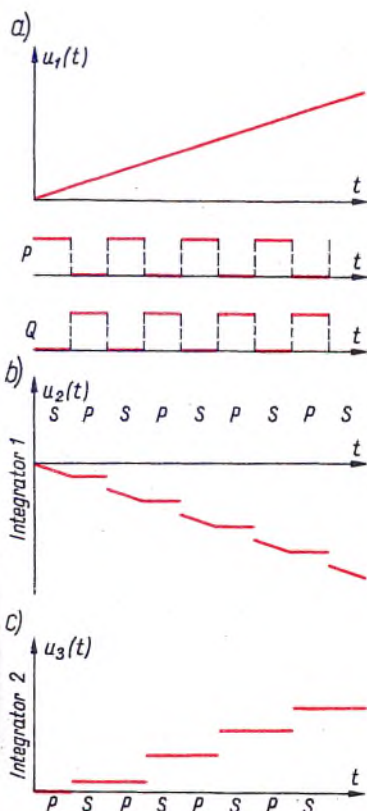
Widać, że programowanie maszyny iteracyjnej jest znacznie bardziej złożone, niż programowanie konwencjonalnej maszyny analogowej. Dużą trudność stanowi opracowanie zasad współpracy integratorów i realizacja układu sterowania.

·7.26· Jaki jest najprostszy program realizowany na maszynie iteracyjnej?

Niewątpliwie najprostszym programem możliwym do zrealizowania na maszynie iteracyjnej jest pamięć analogowa (tzw. para pamięciowa). Pamięć stanowi połączenie dwóch integratorów sterowanych w sposób pokazany na rys. 7.26A. Jak widać, wejścia liczące nie są tu używane, pracują tylko wejścia śledzące. Integratory pracują na przemian, przy czym mogą znajdować się tylko w jednym z dwóch stanów — „śledzenie” lub „pamiętanie”.

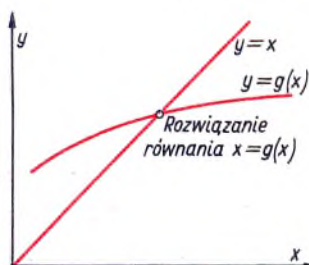


·7.26A· Para pamięciowa; a) schemat układu; b) sygnały sterujące



·7.26B· Ilustracja działania pary pamięciowej

S — stan śledzenia; P — stan pamiętania



·7.27A· Interpretacja graficzna rozwiązania równania $x = g(x)$

Na rysunku 7.26B(c) przedstawiono sygnał otrzymany na wyjściu układu, gdy na wejście podano sygnał liniowo narastający (rys. 7.26B(a)). Jak widać, układ **zapamiętuje** wartość sygnału istniejącego na wejściu w chwili zmiany stanu integratora I ze „śledzenia” na „pamiętanie”, i **pamięta** ten sygnał aż do ponownej zmiany stanu integratora I. Stąd też nazwa układu — para pamięciowa lub pamięć analogowa.

Zauważmy, że pracę pary pamięciowej można opisać za pomocą równania różnicowego

$$y_k = x_{k-1}$$

w którym

y_k — sygnał wyjściowy pary pamięciowej w chwili k ;

x_{k-1} — sygnał wejściowy w chwili $k-1$.

Tak więc para pamięciowa jest dyskretnym analogiem integratora — integrator rozwiązywał najprostsze równanie różniczkowe, natomiast za pomocą opisanego układu można rozwiązać najprostsze równanie różnicowe.

·7.27· Jaki można podać przykład zastosowania pary pamięciowej (programu iteracyjnego)?

Zbadajmy zastosowanie pary pamięciowej do realizacji najprostszego programu iteracyjnego. Program ten jest bardzo prosty, jednak jest podstawą większości złożonych programów realizujących znacznie bardziej skomplikowane zadania.

Założmy, że mamy rozwiązać równanie

$$x = g(x)$$

Interpretację graficzną rozwiązania pokazano na rys. 7.27A. Rozwiązaniem jest punkt przecięcia prostej $y = x$, oraz wykresu funkcji $y = g(x)$. Na ogół niemożliwe jest rozwiązywanie analityczne równania, można jednak rozwiązać je numerycznie.

Założmy, że znamy przybliżenie początkowe rozwiązania x_0 . Konstruujemy proces iteracyjny

$$x_{n+1} = g(x_n)$$

W pewnych warunkach proces ten będzie generował przybliżenia rozwiązania x_n zbliżne do rozwiązania dokładnego. Interpretację graficzną procesu podano na rys. 7.27B.

Proces ten może być wykonywany za pomocą maszyny cyfrowej, możliwe jest również zastosowanie maszyny iteracyjnej.

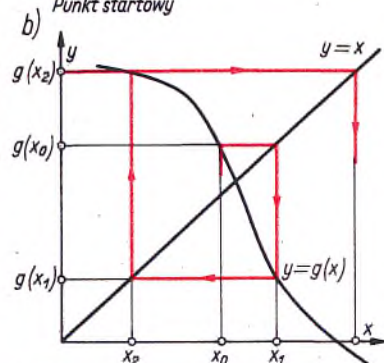
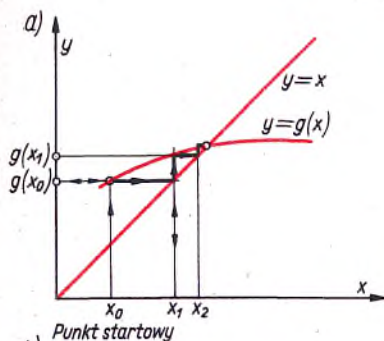
Jeżeli oznaczmy

$$z_n = g(x_n)$$

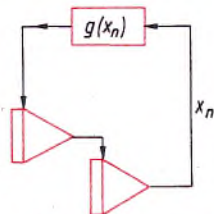
to rozważany proces możemy zapisać

$$x_{n+1} = z_n$$

Równanie takie można rozwiązać stosując parę pamięciową. Wynika stąd natychmiast sposób realizacji procesu ite-



·7.27B· Iteracyjne rozwiązywanie równania $x = g(x)$: a) proces iteracyjny zbieżny; b) proces iteracyjny rozbieżny



·7.27C· Zastosowanie pary pamięciowej do rozwiązywania równania $x = g(x)$

racyjnego (rys. 7.27C). Zauważmy, że nie zakładaliśmy nic o naturze funkcji $g(x_n)$. Funkcja ta może być bardzo złożona — może to być np. układ analogowy obliczający pochodną wskaźnika jakości układu regulacji względem parametrów, np. wzmocnienia. Umożliwia to zastosowanie maszyny iteracyjnej do optymalizacji parametrycznej układu sterowania.

·7.28· Co to jest dwójka akumulacyjna (akumulator)?

Dwójka akumulacyjna jest to układ iteracyjny opisany równaniem różnicowym

$$x_{n+1} = x_n + u_n \quad (7.28.1)$$

w którym u_n — sygnał podawany na wejście dwójki w chwili t_n . Korzystając z analizy pary pamięciowej łatwo podać sposób realizacji dwójki akumulacyjnej (rys. 7.28). Z zależności (7.28.1) wynika że podając na wejście dwójki akumulacyjnej sygnał stały (np. 1) otrzymamy na wyjściu w kolejnych chwilach 0, 1, 2, 3 ... itd. Dwójka akumulacyjna może zatem służyć jako generator funkcji „schodkowej” i najczęściej jest w tym celu stosowana.

·7.29· W jaki sposób można zastosować maszynę iteracyjną do minimalizacji funkcji jednej zmiennej?

Możliwe są dwa sposoby rozwiązania zadania. Pierwszy z nich polega na rozwiązaniu równania

$$g(x) = 0$$

w którym $g(x)$ — pochodna minimalizowanej funkcji.

Równanie to można rozwiązać, stosując parę pamięciową w sposób opisany w pyt. 7.27. Drugi sposób — to zastosowanie dwójki akumulacyjnej do bezpośredniego przeszukiwania funkcji.

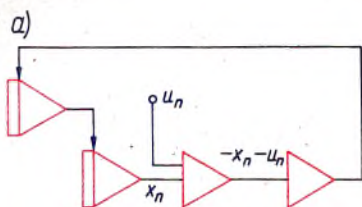
Żałujemy, że na wejście dwójki akumulacyjnej podajemy stałą w czasie, niewielkie napięcie dodatnie. Sygnał wyjściowy dwójki podajemy na wejście układu obliczającego wartość minimalizowanej funkcji, sygnał wejściowy tego układu podajemy na wejście pary pamięciowej. Obie dwójki — pamięciowa i akumulacyjna pracują synchronicznie tzn. są sterowane tymi samymi sygnałami sterującymi (rys. 7.29A). Dwójka akumulacyjna wytwarza przebieg schodkowy, na wyjściu bloku obliczającego wartości funkcji pojawia się napięcie odpowiadające wartości funkcji dla kolejnych przyrostów napięcia wejściowego

$$u_{wy} = f(n \cdot u_0)$$

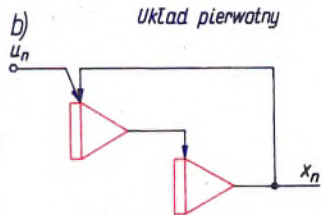
Napięcie na wyjściu pary pamięciowej jest opóźnione o 1 takt, tzn. jest równe

$$\bar{u}_{wy} = f((n-1)u_0)$$

Jeżeli w pewnej chwili pracy układu różnica $\bar{u}_{wy} - u_{wy}$ zmieni znak z dodatniego na ujemny, oznacza to, że napięcie wejściowe przekroczyło punkt, w którym funkcja osiąga mi-

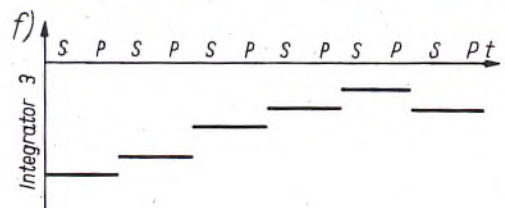
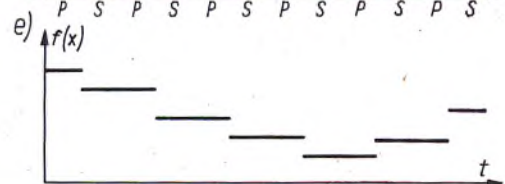
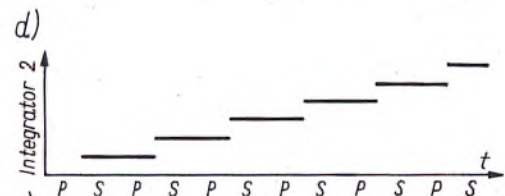
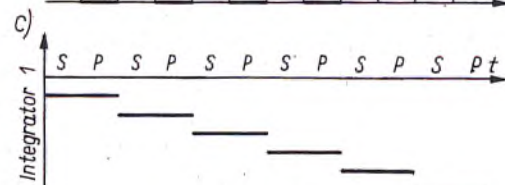
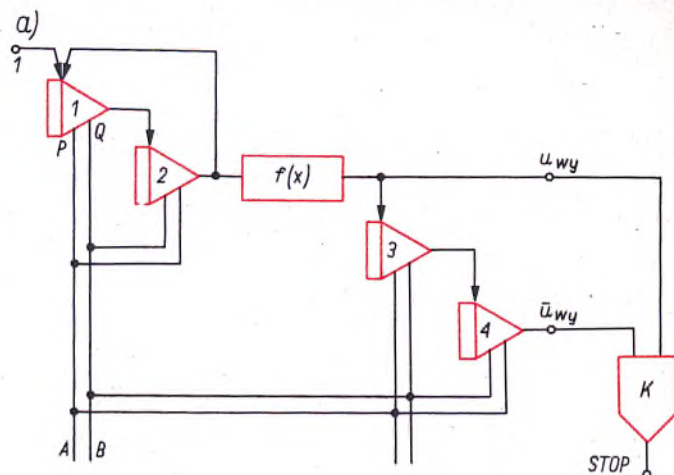


Układ pierwotny



Układ uproszczony

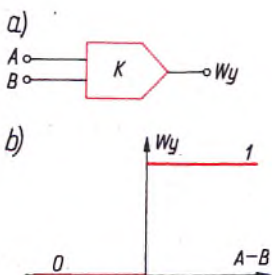
7.28. Sposób realizacji dwójki akumulacyjnej: a) za pomocą integratorów z jednym wejściem śledzącym; b) za pomocą integratorów z dwoma wejściami śledzącymi



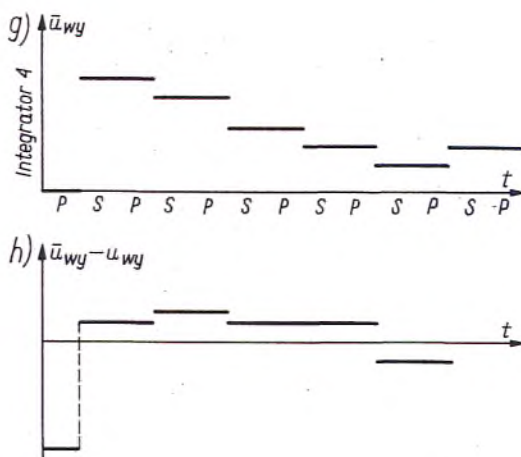
7.29A. Zastosowanie dwójki akumulacyjnej i pary pamięciowej do minimalizacji funkcji metodą przeszukiwania: a) schemat; b...h) przebiegi czasowe;

s — stan śledzenia; p — stan pamiętania

·7.29A· cd.



·7.29B· Komparator; a) symbol; b) charakterystyka

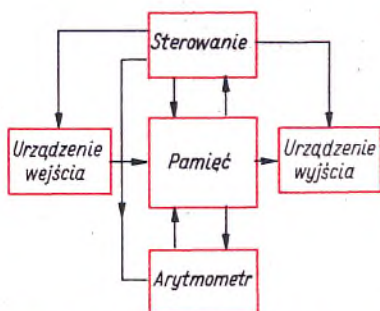


nimum. Należy wówczas wstrzymać pracę urządzenia. Obliczanie różnicy $\bar{u}_{wy} - u_{wy}$ oraz badanie jej znaku można wykonać za pomocą układu zwanego **komparatorem** (rys. 7.29B). Układ ten porównuje napięcia wejściowe — jeżeli różnica $u_A - u_B$ jest dodatnia, na wyjściu układu jest sygnał odpowiadający jedynce logicznej, jeżeli ujemna — zeru logicznemu. Każda maszyna iteracyjna jest wyposażona w gniazdko STOP, podanie jedynki logicznej do tego gniazdka powoduje wstrzymanie pracy maszyny. Schemat układu z komparatorem i przebiegi czasowe podano na rys. 7.29A.

·7.30· Jak jest zbudowana cyfrowa maszyna matematyczna i czym różni się od maszyny analogowej?

Maszyna cyfrowa jest urządzeniem obliczeniowym operującym **liczbami** (wynika to zresztą z samej nazwy). Umożliwia ona wykonywanie 4 podstawowych działań arytmetycznych. Struktura maszyny jest przystosowana do wykonywanych zadań — składa się ona z **arytmometru** wykonującego działania arytmetyczne, **pamięci** mającej zgromadzić odpowiednią ilość liczb, **układu sterowania** organizującego współpracę między poszczególnymi elementami maszyny, oraz **urządzeń wejścia-wyjścia** umożliwiających wprowadzanie danych i odczyt wyników (rys. 7.30). Poszczególne operacje maszyny cyfrowej wykonywane są **kolejno** na podstawie programu zapisanego w pamięci.

Podstawową różnicą między obydwooma typami maszyn jest to, że maszyna analogowa wykonuje działania na **funkcjach czasu** (sygnałach), natomiast maszyna cyfrowa na **liczbach**. Druga różnica wynika ze sposobu działania maszyn oraz programowania. W maszynie analogowej wszystkie operacje są wykonywane **jednocześnie**. Jeżeli program zawiera np. kilka operacji dodawania, to każda z nich jest wykonywana za pomocą innego układu operacyjnego, a więc obliczenia są wykonywane **równolegle**.

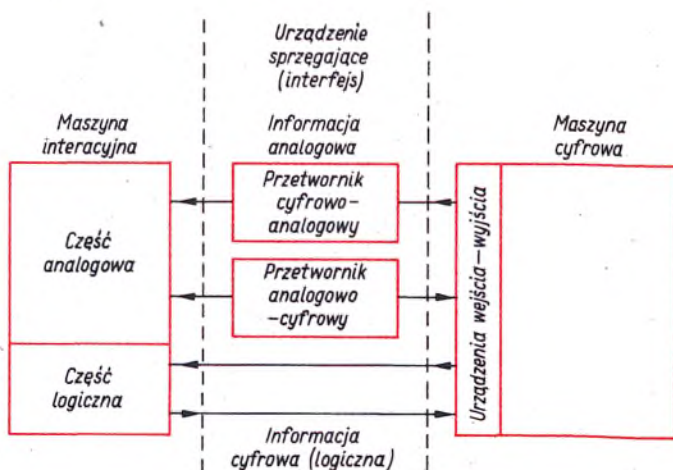


·7.30· Schemat organizacji cyfrowej maszyny matematycznej

W maszynie cyfrowej jest tylko jeden arytmometr; jeżeli jest kilka operacji dodawania, są one wykonywane kolejno, jedna za drugą — obliczenia w maszynie cyfrowej są wykonywane **szeregowo**. Fakt ten wyjaśnia, dlaczego maszyny analogowe są znacznie szybsze niż maszyny cyfrowe (przykład: pewne równanie różniczkowe było rozwiązywane na maszynie analogowej MEDA-TC z szybkością 10 rozwiązań na sekundę; to samo równanie rozwiązywane jednokrotnie na maszynie cyfrowej ODRA 1325 wymagało ok. 30 minut pracy maszyny). Trzecia różnica wynika ze sposobu realizacji elementów operacyjnych maszyny cyfrowej i analogowej. Liczby w maszynie cyfrowej są reprezentowane z dosyć dużą dokładnością (9 do 18 miejsc znaczących). Z zasady działania arytmometru wynika, że utrata dokładności podczas wykonywania jednego działania jest niewielka. W maszynie analogowej sygnały są przetwarzane przez układy, których dokładność jest rzadko większa niż 0,1%. Ponieważ program składa się z wielu takich elementów, dokładność rozwiązania jest około 0,5 ... 1%. Z powyższego wynika, że dokładność maszyny cyfrowej jest większa niż maszyny analogowej. W praktyce jednak przy rozwiązywaniu równań różniczkowych za pomocą maszyny cyfrowej trzeba wykonać tak olbrzymią liczbę obliczeń, że kumulujące się błędy zaokrągleń powodują znaczną utratę dokładności wyniku; niekiedy otrzymuje się wynik zupełnie bez sensu.

•7.31• Co to jest hybrydowa maszyna matematyczna?

Maszyna hybrydowa (rys. 7.31) jest to maszyna iteracyjna sprzęgnięta z maszyną cyfrową. Maszyna cyfrowa ma możliwość sterowania integratorów maszyny analogowej, przesyłania informacji cyfrowej i analogowej do maszyny analogowej. W ten sposób część obliczeń wymagających dużej dokładności (np. rozwiązywanie układów równań algebraicznych) może być wykonywana przez maszynę cyfrową,



•7.31• Schemat organizacji hybrydowej maszyny matematycznej

natomiast rozwiązywanie równań różniczkowych — przez maszynę analogową.

·7.32· Czy za pomocą maszyny cyfrowej można rozwiązać równanie różniczkowe?

Tak, wymaga to jednak pewnego przeformułowania problemu. Rozważmy przykład: Rozwiązać równanie różniczkowe

$$\dot{y}(t) = -y(t) \quad (7.32.1)$$

z warunkiem początkowym

$$y(0) = 1$$

Szukamy **funkcji**, która spełnia to równanie. Jak wiemy, maszyna cyfrowa nie może operować funkcjami, może tylko operować liczbami. Tak więc zamiast szukać funkcji $y(t)$ musimy zadowolić się jej wartościami $y(t_i)$ w odpowiednio wybranych chwilach czasu t_i . Równanie (7.32.1) należy zmodyfikować tak, aby maszyna cyfrowa mogła na jego podstawie wyznaczać wartości $y(t_i)$.

W dalszym ciągu zakładamy będziemy, że punkty t_i rozłożone są równomiernie, a odległość między dwoma kolejnymi punktami jest równa τ .

Najprostszy sposób modyfikacji równania (7.32.1) polega na **dyskretyzacji**, tzn. zastąpieniu pochodnej funkcji $y(t)$ ilorazem różnicowym

$$y'(t_i) = \frac{y(t_{i+1}) - y(t_i)}{\tau}$$

Przybliżenie to jest zadowalające jeżeli τ jest dostatecznie małe (rys. 7.32a). Otrzymujemy równanie zmodyfikowane

$$\frac{y(t_{i+1}) - y(t_i)}{\tau} = -y(t_i)$$

Można napisać w postaci

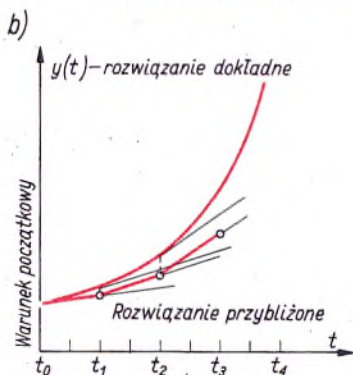
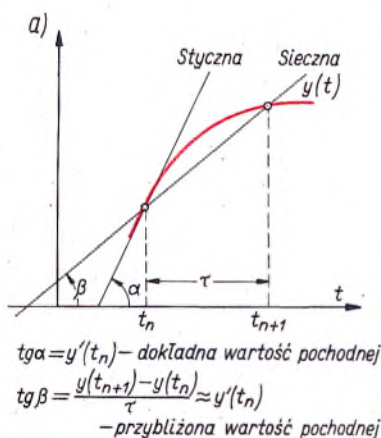
$$y(t_{i+1}) = y(t_i)(1 - \tau) \quad (7.32.2)$$

Równanie to stanowi prostą receptę, według której możemy wyznaczać kolejne wartości rozwiązania — jeżeli znamy

7.32. Rozwiązanie dokładne i przybliżone równania różniczkowego

$$\dot{y}(t) = -y(t), \quad y(0) = 1$$

t	Rozwiązanie dokładne	Rozwiązanie przybliżone	
		$\tau = 0,01$	$\tau = 0,1$
0	1,0	1,0	1,0
0,2	0,81873	0,81791	0,81
0,4	0,67032	0,66897	0,65609
0,6	0,54881	0,54716	0,53144
0,8	0,44933	0,44752	0,43047
1,0	0,36788	0,36603	0,34868



·7.32· Zamiana równania różniczkowego na równanie różnicowe przez aproksymację pochodnej różnicą skończoną

$y(t_i)$, to że wzoru (7.32.2) wynika natychmiast wartość rozwiązania w następnej chwili (rys. 7.32b). W tablicy 7.32 przedstawiono wyniki rozwiązania równania (7.32.2) oraz wartość rozwiązania dokładnego otrzymanego z zależności $y(t) = e^{-t}$

Obliczenie prowadzono przy $\tau = 0,01$

·7.33· Czy rozwiązanie równania różniczkowego wyznaczone za pomocą maszyny cyfrowej jest zawsze dobrym przybliżeniem rozwiązania dokładnego?

Nie. Rozważmy równanie (7.32.2) z poprzedniego pytania. Można łatwo znaleźć jego rozwiązanie.

Jeżeli $y(t_0) = 1$ (warunek początkowy), to

$$y(t_i) = (1 - \tau)^i$$

Rozwiązanie dokładne równania (7.32.1) ma postać

$$\tilde{y}(t) = e^{-t}$$

Wartość tego rozwiązania w chwili $t_i = i\tau$ jest równa

$$y(t_i) = e^{-i\tau} = (e^{-\tau})^i$$

Porównując oba wzory widzimy, że w rozwiązaniu przybliżonym wyrażenie $e^{-\tau}$ jest zastąpione przez $1 - \tau$. Jeżeli τ będzie dobrane w sposób niewłaściwy, różnica między wartością rozwiązania dokładnego i przybliżonego może być bardzo znaczna. Istotnie, jeżeli przyjmiemy np. $\tau = 3$, wówczas rozwiązanie równania przybliżonego

$$y(t_i) = (-2)^i$$

Oczywiście rozwiązanie to nie ma nic wspólnego z rozwiązaniem dokładnym. Opisane zjawisko nazywamy niestabilnością **metody numerycznej**. Należy zwrócić uwagę, że niestabilność metody nie ma nic wspólnego ze stabilnością samego równania różniczkowego. Przebieg rozwiązań przybliżonych dla $\tau = 1,5$; $0,8$; $0,1$ pokazano na rys. 7.33.

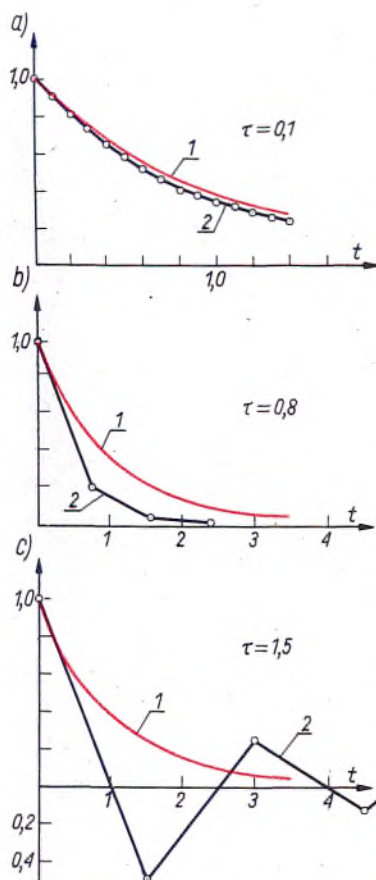
Tak więc wartość τ należy dobierać tak, aby metoda numeryczna była stabilna. W naszym przypadku powinna być spełniona nierówność

$$0 < \tau < 1$$

Spełnienie tego warunku nie wystarczy jednak do otrzymania prawidłowego rozwiązania. Wybierając np. $\tau = 0,95$ uzyskamy co prawda metodę stabilną, lecz dokładność rozwiązania będzie bardzo mała. Tak więc wartość τ należy wybierać kierując się dwoma kryteriami — stabilnością metody różnicowej, oraz dokładnością otrzymanego rozwiązania.

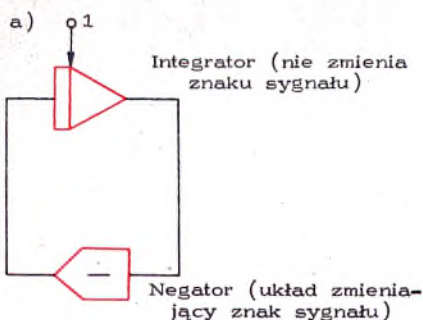
·7.34· Co to jest symulator cyfrowy maszyny analogowej?

Programowanie maszyny cyfrowej jest dosyć pracochłonne i wymaga znacznych umiejętności. Programowanie maszyny analogowej jest znacznie prostsze — samo równanie na-

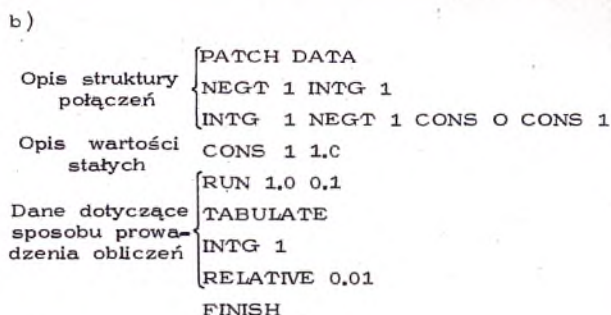


·7.33· Wpływ kroku dyskretyzacji τ na dokładność numerycznego rozwiązania równania różniczkowego: a) krok τ dobrany właściwie; b, c) metoda stabilna, lecz krok τ źle dobrany ze względu na dokładność;

1 — rozwiązanie dokładne; 2 — rozwiązanie przybliżone



7.34 Program do rozwiązywania równania $\dot{y} = -y$ za pomocą symulatora cyfrowego maszyny analogowej



rzuca strukturę programu. Jednocześnie maszyny cyfrowe są znacznie bardziej rozpowszechnione, niż maszyny analogowe. Tak więc zaproponowano system w którym program ułożony dla maszyny analogowej wprowadza się do maszyny cyfrowej, ta z kolei tłumaczy ten program do postaci umożliwiającej wykonywanie obliczeń na maszynie cyfrowej (wykonuje operacje dyskretyzacji) i rozwiązuje zadanie. System taki nazywa się cyfrowym symulatorem maszyny analogowej. Umożliwia on znaczne uproszczenie rozwiązywania równań różniczkowych za pomocą maszyny cyfrowej. Na rysunku 7.34 przedstawiono program do rozwiązywania równania z pyt. 7.32.

